ARO 26394.4-MA

2

ITERATIVE PROCEDURES FOR EXACT MAXIMUM LIKELIHOOD ESTIMATION IN THE FIRST-ORDER GAUSSIAN MOVING AVERAGE MODEL

T. W. Anderson and R. P. Mentz
Stanford University and University of Tucumán

TECHNICAL REPORT NO. 25
NOVEMBER 1990

U. S. Army Research Office
Contract DAAL03-89-K-0033
Theodore W. Anderson, Project Director

Department of Statistics
Stanford University
Stanford, California

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER ARO 26394.4-MA	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) Iterative Procedures for Exact Maximum Likelihood Estimation in the First-order Gaussian Moving Average Model		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T. W. Anderson & R. P. Mentz		8. CONTRACT OR GRANT NUMBER(s) DAAL03 89 K 0033
9. PERFORMING ORGANIZATION NAME AND ADDRESS Stanford University Dept. of Statistics - Sequoia Hall Stanford, CA 94305-4065		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE November, 1990 13. NUMBER OF PAGES 57
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Maximum likelihood, first-order moving average model, iterative procedures

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See reverse side for abstract.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

Estimation of the parameters of a first-order Gaussian moving average model is treated in detail. Iterative methods in both the time and "frequency" domains are based on the maximization of the exact likelihood. Several methods for evaluating the necessary quadratic forms and traces are presented. The procedures are compared with each other and with alternative procedures.

Contents

1. Introduction

Maximum likelihood estimates of the parameters in Gaussian time series maximize the likelihood function

$$L(\boldsymbol{\theta}) = (2\pi)^{-T/2}|\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\right\}, \tag{1.1}$$

where $\boldsymbol{y}$ is the vector of observations, $\mathcal{E}\boldsymbol{y} = \mathbf{0}$, $\mathcal{E}\boldsymbol{y}\boldsymbol{y}' = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. Exact maximum likelihood estimation is complicated because even for standard models the needed determinants and inverses are either not known in closed forms or, if they are known, they involve complicated parametric functions. Thus only for the first-order autoregressive model are the explicitly forms of the maximum likelihood estimators of its two parameters known (Hasza, 1980).

One procedure that is often used is to operate mathematically with the likelihood function, so that it can be evaluated numerically at any point of the parameter space, and then to optimize the function by varying values of the parameters, using an efficient computer program. For example, the IMSL Library (1979) package of Fortran subroutines uses a "modified steepest descent algorithm" to compute estimates of the parameters of ARIMA models; the statistical package BMDP (1985) uses "Gauss-Marquardt methods" to perform linear and nonlinear estimations. One idea in this area by Box and Jenkins (1976) was the "backcasting" procedure to evaluate the approximate likelihood function in some time series models. Useful suggestions have been the Cholesky decomposition of the covariance matrix and "Woodbury's formula," as in Phadke and Kedem (1978), for example. Ansley (1979) studied several approaches and proposed a new algorithm; he reviewed earlier work by Newbold (1974), Dent (1977), Ali (1977), Osborn (1977), and Hillmer and Tiao (1979). He showed the equivalence of the proposals by Newbold (1974) and Dent (1977), and related this approach to Ali (1977). See also Nicholls and Hall (1979).

Another approach is to derive an iterative procedure. Some iterative procedures use the likelihood equations deduced by setting the derivatives of the likelihood function equal to 0 to obtain a procedure of the form $\widehat{\boldsymbol{\theta}}_i = g(\widehat{\boldsymbol{\theta}}_{i-1})$ for some function g depending on $\boldsymbol{y}$, where $\widehat{\boldsymbol{\theta}}_i$, $i = 1, 2, \ldots$ are successive numerical values of the estimates, with $\widehat{\boldsymbol{\theta}}_0$ as a starting value. Anderson (1975, 1977) considered several procedures for exact maximum likelihood estimation, and some approximations. Godolphin and de Gooijer (1982) presented a procedure for the first-order moving average model. In general, deriving iterative procedures tends to involve more mathematical elaborations of, and more knowledge about the likelihood function, than the approaches mentioned in the previous paragraph. A general

iterative procedure for estimation purposes is the so-called "EM-algorithm" (Dempster, et al., 1977), proposed initially for some missing value statistical problems; in our context the initial values can be thought of as missing values, and when applicable the EM algorithm will suggest a sequence of conditional expectation-optimization steps for the iterative computation of the maximum likelihood estimates. Another approach involves Kalman filtering techniques; see, for example, Harvey (1981). The connection of the present work with the EM-algorithm and with Kalman filtering will be considered briefly in Section 8.

The purpose of this paper is to consider in detail iterative procedures for exact maximum likelihood estimation in the first-order Gaussian moving average model. Section 2 introduces the model; Section 3 deals in general with iterative procedures, and the specific procedures are derived in Section 4. Sections 5, 6, and 7 contain the evaluations of quadratic forms and traces in the time and frequency domains. Section 8 contains various comments on the procedures. Section 9 contains evaluations of the numbers of operations needed to compute certain traces and quadratic forms. Finally Section 10 considers the approach of Box and Jenkins (1976) to the estimation problem.

2. The First-Order Moving Average Model

The Gaussian moving average process of order 1 with mean 0, denoted here by MA(1), is defined by

$$y_t = u_t + \alpha u_{t-1}, \quad t = \ldots, -1, 0, 1, \ldots , \tag{2.1}$$

where the y_t are observable, the u_t are unobservable independent normal random variables with $\mathcal{E} u_t = 0$, $\mathcal{E} u_t^2 = \sigma^2$, $0 < \sigma^2 < \infty$, and α and σ^2 are parameters. The Gaussian moving average is a stationary stochastic process for any value of α.

The autocovariance (or covariance) sequence of the process is

$$\begin{aligned} \sigma_s &= \sigma^2(1+\alpha^2), && s = 0, \\ &= \sigma^2 \alpha, && |s| = 1, \\ &= 0, && |s| > 1. \end{aligned} \tag{2.2}$$

The autocorrelation (or correlation) sequence of the process is

$$\begin{aligned} \rho_s &= 1, && s = 0, \\ &= \tfrac{\alpha}{1+\alpha^2}, && |s| = 1, \\ &= 0, && |s| > 1. \end{aligned} \tag{2.3}$$

For convenience, we write ρ for ρ_1. The covariance sequence satisfies the Fourier inversion formula

$$\sigma_s = \int_{-\pi}^{\pi} e^{i\lambda s} f(\lambda) d\lambda, \tag{2.4}$$

where $f(\lambda)$ is the spectral density of the process, given by

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| e^{i\lambda} + \alpha \right|^2 = \frac{\sigma^2}{2\pi}(1 + \alpha^2 + 2\alpha \cos \lambda), \quad -\pi \le \lambda \le \pi. \tag{2.5}$$

If $y_1, \ldots, y_T$ is a sample from (2.1), $\boldsymbol{y} = (y_1, \ldots, y_T)'$ has a multivariate normal distribution with expectation $\mathcal{E}\boldsymbol{y} = \boldsymbol{0}$ and covariance matrix $\mathcal{E}\boldsymbol{y}\boldsymbol{y}' = \boldsymbol{\Sigma} = (\sigma_{|i-j|})$.

Let us introduce the $T \times T$ matrices $\boldsymbol{P}$ and $\boldsymbol{R}$ (the correlation matrix) defined by

$$\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{P} = \sigma^2(1 + \alpha^2)\boldsymbol{R} = \sigma_0 \boldsymbol{R}. \tag{2.6}$$

We note that $\boldsymbol{\Sigma}$, $\boldsymbol{P}$, and $\boldsymbol{R}$ can be written as linear combinations of the identity matrix $\boldsymbol{I}$ and the matrix $\boldsymbol{G}$ that is symmetric, has 1's along the diagonals immediately above and below the main diagonal, and 0's elsewhere. In this notation

$$\boldsymbol{\Sigma} = \sigma_0 \boldsymbol{I} + \sigma_1 \boldsymbol{G} = \sigma^2(1 + \alpha^2)\boldsymbol{I} + \sigma^2 \alpha \boldsymbol{G}, \tag{2.7}$$

$$\boldsymbol{P} = (1 + \alpha^2)\boldsymbol{I} + \alpha \boldsymbol{G}, \tag{2.8}$$

$$\boldsymbol{R} = \boldsymbol{I} + \frac{\alpha}{1 + \alpha^2}\boldsymbol{G} = \boldsymbol{I} + \rho \boldsymbol{G}. \tag{2.9}$$

For any α and T, $\boldsymbol{\Sigma}$, $\boldsymbol{P}$, and $\boldsymbol{R}$ are positive definite, the first because we also assume that $\sigma^2 > 0$. As a function of ρ, $\boldsymbol{R}$ is positive definite for $-a < \rho < a$, where $a = 1/\{2\cos[\pi/(T+1)]\}$; see Anderson and Takemura (1986).

The likelihood function can be written as a function of α and σ^2 as

$$L^*(\alpha, \sigma^2) = (2\pi)^{-T/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} \boldsymbol{y}' \boldsymbol{\Sigma}^{-1} \boldsymbol{y} \right\} \tag{2.10}$$

$$= (2\pi)^{-T/2} (\sigma^2)^{-T/2} |\boldsymbol{P}|^{-1/2} \exp\left\{ -\frac{1}{2\sigma^2} \boldsymbol{y}' \boldsymbol{P}^{-1} \boldsymbol{y} \right\}. \tag{2.11}$$

Instead of operating with (2.11) for purposes of maximum likelihood estimation, we can separate the analysis into two parts: we maximize (2.11) with respect to σ^2 at

$$\widehat{\sigma}^2 = \frac{1}{T} \boldsymbol{y}' \boldsymbol{P}^{-1} \boldsymbol{y}, \tag{2.12}$$

and then substitute $\sigma^2 = \widehat{\sigma}^2$ in (2.11) to derive the "concentrated likelihood function," which is a constant times the square root of

$$n^*(\alpha) = \frac{1}{|\boldsymbol{P}|(\boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y})^T}. \tag{2.13}$$

Since $\rho = \alpha/(1+\alpha^2) = (1/\alpha)/[1+(1/\alpha)^2]$, the likelihood function attains all possible values on the set for which $|\alpha| \leq 1$, $0 < \sigma^2 < \infty$. Hence, without loss of generality we restrict attention to this set. Note that $|\alpha| < 1$ is the condition for invertibility, that is, to express (2.1) as an infinite autoregression.

In terms of σ_0 and ρ the likelihood function can be written as

$$L(\sigma_0, \rho) = (2\pi)^{-T/2}(\sigma_0)^{-T/2}|\boldsymbol{R}|^{-1/2} \exp\left\{-\frac{1}{2\sigma_0}\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}\right\}. \tag{2.14}$$

If the analysis is separated into two parts, we maximize (2.14) with respect to σ_0 at

$$\widehat{\sigma}_0 = \frac{1}{T}\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}, \tag{2.15}$$

and then maximize with respect to ρ the function

$$n(\rho) = \frac{1}{|\boldsymbol{R}|(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y})^T}. \tag{2.16}$$

Note that $n^*(\alpha) = n[\rho(\alpha)]$, where $\rho(\alpha) = \alpha/(1+\alpha^2)$.

For $|\alpha| \leq 1$ and $|\rho| \leq \frac{1}{2}$ the following three pairs provide alternative equivalent parametrizations for (2.1): α and σ^2, σ_0 and σ_1, ρ and σ_0. For example,

$$\alpha = \frac{1-\sqrt{1-4(\sigma_1/\sigma_0)^2}}{2(\sigma_1/\sigma_0)} = \frac{1-\sqrt{1-4\rho^2}}{2\rho}. \tag{2.17}$$

Hence, for purposes of maximum likelihood estimation we can operate with (2.10) as a function of σ_0 and σ_1, with (2.11) as a function of α and σ^2, or with (2.14) as a function of ρ and σ_0. Similarly, in (2.13) we operate with a function of α, and in (2.16) with a function of ρ. The relationship between the two parametrizations will be studied in more detail in Section 3.

For further details about the moving average model see, for example, Anderson (1971) or Anderson and Mentz (1980).

3. Some Approaches to the Iterative Estimation by Maximum Likelihood

The method of maximum likelihood proposes to estimate the parameters by maximizing (2.10), (2.11) or (2.14); alternatively, we can use (2.12) and (2.13), or (2.15) and (2.16). A basic difficulty comes from the complicated nature of the parametric functions that are involved. Let $\Delta_T = |\boldsymbol{P}|$; then as functions of α and ρ we have for $|\alpha| < 1$,

$$\Delta_T = \frac{1-\alpha^{2(T+1)}}{1-\alpha^2} \tag{3.1}$$

$$= \frac{1}{2^T} \sum_{j=0}^{[T/2]} \binom{T+1}{2j+1} (1-4\rho^2)^j, \quad T = 0, 1, \ldots, \tag{3.2}$$

respectively. These results can be verified by showing that Δ_T satisfies the homogeneous difference equation

$$\Delta_T - (1+\alpha^2)\Delta_{T-1} + \alpha^2 \Delta_{T-2} = 0, \quad T = 2, 3, \ldots,$$

with $\Delta_0 = 1$, $\Delta_1 = (1-\alpha^4)/(1-\alpha^2)$. See Anderson (1971), Section 6.7, for example. If p^{ij} denotes the i,j-th component of $\boldsymbol{P}^{-1}$, then

$$p^{ij} = (-\alpha)^{j-i} \frac{\Delta_{i-1}\Delta_{T-j}}{\Delta_T}, \qquad j \geq i, \tag{3.3}$$

so that as a function of α,

$$p^{ij} = (-\alpha)^{j-i} \frac{(1-\alpha^{2i})(1-\alpha^{2(T-j+1)})}{(1-\alpha^2)(1-\alpha^{2(T+1)})}, \qquad j \geq i. \tag{3.4}$$

See Shaman (1969).

3.1. Four aspects of maximum likelihood estimation

I. Likelihood vs. concentrated likelihood functions

As indicated in Section 2, operating with (2.10) or (2.11) in terms of α and σ^2 is mathematically equivalent to operating with (2.12) and with the concentrated likelihood function or with (2.13); similarly, operating with (2.14) in terms of σ_0 and ρ is mathematically equivalent to operating with (2.15) and (2.16), in the sense that the solutions are the same. However, for the same set of parameters, the estimating equations are not the same for the likelihood and concentrated likelihood functions.

II. α and σ^2 vs. ρ and σ_0

We have a choice of parameters to consider. As indicated in Section 2, the three pairs that we discussed provide equivalent alternative parametrizations for the model (2.1). Since the maximum likelihood estimation procedure is invariant under such type of transformations of the parameters it follows that from a mathematical point of view it is immaterial with which set we choose to operate. However, different sets of parameters may lead to different estimating equations.

III. Time vs. frequency domains

We presented our results so far (except for (2.4) and (2.5)) in the time domain. We can consider the effect of a Fourier transformation. Let $\boldsymbol{K}$ be the orthogonal $T \times T$ matrix with components

$$\sqrt{\frac{2}{T+1}} \sin \frac{\pi j k}{T+1}, \qquad j, k = 1, \ldots, T, \tag{3.5}$$

and let $\boldsymbol{D}$ be the diagonal matrix with diagonal elements

$$d_j = 2 \cos \frac{\pi j}{T+1}, \qquad j = 1, \ldots, T. \tag{3.6}$$

Then

$$\boldsymbol{K}'\boldsymbol{K} = \boldsymbol{K}\boldsymbol{K}' = \boldsymbol{I}, \qquad \boldsymbol{K}'\boldsymbol{G}\boldsymbol{K} = \boldsymbol{D}. \tag{3.7}$$

We have a way to diagonalize the matrix $\boldsymbol{G}$ appearing in (2.7), (2.8), and (2.9). If $\boldsymbol{y} = \boldsymbol{K}\boldsymbol{z}$, then $\boldsymbol{z} = \boldsymbol{K}'\boldsymbol{y}$, that is,

$$z_j = \sqrt{\frac{2}{T+1}} \sum_{k=1}^{T} y_k \sin \frac{\pi j k}{T+1}, \qquad j = 1, \ldots, T; \tag{3.8}$$

then $\boldsymbol{z}$ is multivariate normal with $\mathcal{E}\boldsymbol{z} = \boldsymbol{0}$ and

$$\mathcal{E}\boldsymbol{z}\boldsymbol{z}' = \sigma_0 \boldsymbol{I} + \sigma_1 \boldsymbol{D} = \sigma_0(\boldsymbol{I} + \rho \boldsymbol{D}). \tag{3.9}$$

This then provides an alternative approach that we may call a "frequency domain approach": any expression in terms of $\boldsymbol{G}$ can be translated into an expression in terms of $\boldsymbol{D}$, and any method formally presented in terms of $\boldsymbol{y}$ can be translated into a method presented in terms of $\boldsymbol{z}$.

IV. Scoring vs. Newton-Raphson

To maximize the L or n functions (or the L^* or n^* functions) we have available two procedures based on a Taylor's expansion. We illustrate this with

$$\log n(\rho) = -\log|\boldsymbol{R}| - T\log(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}) \tag{3.10}$$

coming from (2.16). The expansion of its derivative with respect to ρ around a value ρ_0 is

$$\frac{d}{d\rho}\log n(\rho) = \frac{d}{d\rho}\log n(\rho)\bigg|_{\rho=\rho_0} + (\rho-\rho_0)\frac{d^2}{d\rho^2}\log n(\rho)\bigg|_{\rho=\rho_0} + R(\rho,\rho_0), \tag{3.11}$$

where $R(\rho,\rho_0)$ is a remainder. The estimating equation is obtained by setting this derivative equal to 0.

The Newton-Raphson procedure consists in replacing the remainder by 0 and setting $\rho = \widehat{\rho}^{(i)}$ and $\rho_0 = \widehat{\rho}^{(i-1)}$. The iterative procedure is then

$$\left\{-\frac{d^2}{d\rho^2}\log n(\rho)\right\}\widehat{\rho}^{(i)} = \frac{d}{d\rho}\log n(\rho) + \left\{-\frac{d^2}{d\rho^2}\log n(\rho)\right\}\widehat{\rho}^{(i-1)}, \tag{3.12}$$

where all derivatives are evaluated at $\rho = \widehat{\rho}^{(i-1)}$.

In the method of scoring the second derivative is replaced by its expectation, where the random vector $\boldsymbol{y}$ is taken with distribution having parameter $\rho = \rho_0$.

About these alternative approaches we note that the dichotomies likelihood vs. concentrated likelihood functions and scoring vs. Newton-Raphson procedures, arise from theoretical considerations. The other dichotomies, α and σ^2 vs. ρ and σ_0 and time vs. frequency domains, are motivated on computational grounds.

From this analysis it follows that to estimate the parameters of the model (2.1) by maximum likelihood under normality, we have sixteen alternative approaches, which may lead to different iterative procedures. Some of these have already been presented in the literature, as will be noted in Section 4.

Anderson (1977) emphasized these dichotomies, while operating with the likelihood function.

3.2. Relations between two parametrizations

Since $\sigma_0 = (1+\alpha^2)\sigma^2 = \sigma_0(\alpha,\sigma^2)$, say, and $\rho = \alpha/(1+\alpha^2) = \rho(\alpha)$, say,

$$L^*(\alpha,\sigma^2) = L[\sigma_0(\alpha,\sigma^2), \rho(\alpha)]. \tag{3.13}$$

In $L(\sigma_0, \rho)$ the range of ρ is $-a < \rho < a$, where $a = 1/\{2\cos[\pi/(T+1)]\}$. A maximum of $L(\sigma_0, \rho)$ occurs at $\widehat{\sigma}_0 > 0$ and $-a < \widehat{\rho} < a$ (with probability 1). These values satisfy the likelihood equations

$$\frac{\partial \log L(\sigma_0, \rho)}{\partial \sigma_0} = 0, \tag{3.14}$$

$$\frac{\partial \log L(\sigma_0, \rho)}{\partial \rho} = 0. \tag{3.15}$$

For $T = 2$ the solution to (3.14), (3.15) is unique, but for $T > 2$ there may be multiple solutions to (3.14), (3.15). The maximum likelihood estimates $\widehat{\sigma}_0, \widehat{\rho}$ are the solution that makes $L(\sigma_0, \rho)$ largest within the range $\sigma_0 > 0$, $-a < \rho < a$.

In $L^*(\alpha, \sigma^2)$ the range of α is $-1 \leq \alpha \leq 1$, and (with probability 1) the maximum occurs at $-1 \leq \widehat{\alpha} \leq 1$ and $\widehat{\sigma}^2 > 0$. These values satisfy the derivative equations

$$\begin{aligned} \frac{\partial \log L^*(\alpha, \sigma^2)}{\partial \alpha} &= \frac{\partial \log L(\sigma_0, \rho)}{\partial \sigma_0} \frac{\partial \sigma_0}{\partial \alpha} + \frac{\partial \log L(\sigma_0, \rho)}{\partial \rho} \frac{d\rho}{d\alpha} \\ &= \frac{\partial \log L(\sigma_0, \rho)}{\partial \sigma_0} 2\alpha\sigma^2 + \frac{\partial \log L(\sigma_0, \rho)}{\partial \rho} \frac{1-\alpha^2}{(1+\alpha^2)^2} = 0, \end{aligned} \tag{3.16}$$

$$\begin{aligned} \frac{\partial \log L^*(\alpha, \sigma^2)}{\partial \sigma^2} &= \frac{\partial \log L(\sigma_0, \rho)}{\partial \sigma_0} \frac{\partial \sigma_0}{\partial \sigma^2} \\ &= \frac{\partial \log L(\sigma_0, \rho)}{\partial \sigma_0} (1+\alpha^2) = 0. \end{aligned} \tag{3.17}$$

If (3.14) is solved for $\sigma_0 = \widehat{\sigma}_0(\rho)$ as given in (2.15) and substituted back into (3.15), we obtain $\frac{1}{2}$ times

$$\frac{d \log n(\rho)}{d\rho} = -\operatorname{tr} \boldsymbol{R}^{-1}\boldsymbol{G} + T\, \frac{\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}} = 0. \tag{3.18}$$

If (3.17) is solved for $\sigma^2 = \widehat{\sigma}^2(\alpha)$ and substituted back into (3.16), we obtain $\frac{1}{2}$ times

$$\frac{d \log n^*(\alpha)}{d\alpha} = \left(-\operatorname{tr} \boldsymbol{R}^{-1}\boldsymbol{G} + T \frac{\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}}\right) \frac{1-\alpha^2}{(1+\alpha^2)^2} = 0. \tag{3.19}$$

where $\boldsymbol{R} = \boldsymbol{I} + \rho(\alpha)\boldsymbol{G}$. Note that (3.19) has the solutions $\alpha = 1$ and -1. Any other solution α^* to (3.19) yields a solution $\rho^* = \rho(\alpha^*)$ to (3.18). However, a solution ρ^* to (3.18) yields a solution to (3.19) only if $-\frac{1}{2} \leq \rho^* \leq \frac{1}{2}$ because then $\rho^* = \rho(\alpha)$ can be solved for the real root

$$\alpha^* = \frac{1 - \sqrt{1 - 4\rho^{*2}}}{2\rho},$$

which lies in the interval $[-1,1]$.

Consider

$$n(\rho) = \frac{1}{|\boldsymbol{R}|(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y})^T} = \frac{1}{\prod_{t=1}^T (1+\rho d_t)\left(\sum_{t=1}^T \frac{z_t^2}{1+\rho d_t}\right)^T} \tag{3.20}$$

$$= \frac{\left[\prod_{t=1}^T (1+\rho d_t)\right]^{T-1}}{\left[\sum_{t=1}^T z_t^2 \prod_{s\neq t}(1+\rho d_s)\right]^T}.$$

As $\rho \to -1/d_T = 1/d_1 = a$, the numerator approaches 0 and the denominator approaches z_T^{2T}; thus $n(\rho) \to 0$ as $\rho \to \pm a$. The derivative of log $n(\rho)$ is

$$\frac{d \log n(\rho)}{d\rho} = -\sum_{t=1}^T \frac{d_t}{1+\rho d_t} + \frac{T}{\sum_{t=1}^T \frac{z_t^2}{1+\rho d_t}} \sum_{t=1}^T \frac{d_t z_t^2}{(1+\rho d_t)^2}. \tag{3.21}$$

Setting (3.21) to 0 yields

$$T\sum_{t=1}^T \frac{d_t z_t^2}{(1+\rho d_t)^2} - \sum_{s=1}^T \frac{d_s}{1+\rho d_s} \sum_{t=1}^T \frac{z_t^2}{1+\rho d_t} = 0. \tag{3.22}$$

Multiplication by $|\boldsymbol{R}|^2 = \prod_{t=1}^T (1+\rho d_t)^2$ gives

$$T\sum_{t=1}^T z_t^2 d_t \prod_{r\neq t}(1+\rho d_r)^2 - \sum_{s=1}^T d_s \prod_{q\neq s}(1+\rho d_q) \sum_{t=1}^T z_t^2 \prod_{r\neq t}(1+\rho d_r) = 0. \tag{3.23}$$

Note $|R|^2$ is of degree $2T-2$ if T is odd and $2T$ if T is even (because $d_{(T+1)/2} = 0$ for T odd).

If T is even, the first polynomial in (3.23) is

$$(3.24)\; T\sum_{t=1}^T z_t^2 d_t \prod_{r\neq t}(1+\rho d_r)^2$$

$$= T\sum_{t=1}^{\frac{T}{2}} \left[z_t^2(1-\rho d_t)^2 - z_{T+1-t}^2(1+\rho d_t)^2\right] d_t \prod_{\substack{r=1\\ r\neq t}}^{\frac{T}{2}} (1-\rho^2 d_r^2)^2$$

$$= T\sum_{t=1}^{\frac{T}{2}} \left[z_t^2 - z_{T+1-t}^2 - 2\rho d_t(z_t^2 + z_{T+1-t}^2) + (z_t^2 - z_{T+1-t}^2)\rho^2 d_r^2\right] d_t \prod_{\substack{r=1\\ r\neq t}}^{\frac{T}{2}} (1-\rho^2 d_r^2)^2$$

$$= T\sum_{t=1}^{\frac{T}{2}} \frac{z_t^2 - z_{T+1-t}^2}{d_t}\; |\boldsymbol{G}|^2\; \rho^{2(T-1)} - 2T\sum_{t=1}^{\frac{T}{2}} \frac{z_t^2 + z_{T+1-t}^2}{d_r^2}\; |\boldsymbol{G}|^2\; \rho^{2T-3} + \dots$$

$$= T\left\{\sum_{t=1}^T \frac{z_t^2}{d_t}\; \rho^{2T-2} - 2\sum_{t=1}^T \frac{z_t^2}{d_t^2}\; \rho^{2T-3} + \dots\right\}$$

since $d_t \neq 0,\ t = 1, \ldots, T$. For T even, $|G| = (-1)^{T/2}$. If T is odd, the first polynomial is

$$(3.25)\qquad T\sum_{t=1}^{\frac{T-1}{2}}\left[z_t^2(1-\rho d_t)^2 - z_{T+1-t}^2(1+\rho d_t)^2\right]d_t \prod_{\substack{r=1\\ r\neq t}}^{\frac{T-1}{2}}(1-\rho^2 d_r^2)^2$$

$$= T\sum_{t=1}^{\frac{T-1}{2}}\Big[(z_t^2 - z_{T+1-t}^2) - 2\rho d_t(z_t^2 + z_{T+1-t}^2)$$

$$+ \rho^2 d_t^2(z_t^2 - z_{T+1-t}^2)\Big]d_t \prod_{\substack{r=1\\ r\neq t}}^{\frac{T-1}{2}}(1-\rho^2 d_r^2)^2$$

$$= T\rho^{2(T-2)}\sum_{t=1}^{\frac{T-1}{2}}(z_t^2 - z_{T+1-t}^2)d_t^3 \prod_{\substack{r=1\\ r\neq t}}^{\frac{T-1}{2}} d_r^4 + \ldots$$

$$= \rho^{2T-4}T\sum_{\substack{t=1\\ t\neq\frac{T+1}{2}}}^{T}\frac{z_t^2}{d_t}\prod_{\substack{r=1\\ r\neq\frac{T+1}{2}}}^{T} d_r^2 + \ldots$$

$$= \rho^{2T-4}T\sum_{\substack{t=1\\ t\neq\frac{T+1}{2}}}^{T}\frac{z_t^2}{d_t}\left(\frac{T+1}{2}\right)^2 + \ldots\ .$$

If T is even, the first factor in the second term is

$$(3.26)\qquad \sum_{s=1}^{T} d_s \prod_{q\neq s}(1+\rho d_q) = \sum_{s=1}^{\frac{T}{2}}\left[d_s(1-\rho d_s) - d_s(1+\rho d_s)\right]\prod_{\substack{q=1\\ q\neq s}}^{\frac{T}{2}}(1-\rho^2 d_q^2)$$

$$= -2\sum_{s=1}^{\frac{T}{2}}\rho d_s^2\prod_{\substack{q=1\\ q\neq s}}^{\frac{T}{2}}(1-\rho^2 d_q^2)$$

$$= T|G|\rho^{T-1} + \text{coefficient } \times \rho^{T-3} + \ldots$$

$$= (-1)^{\frac{T}{2}}T\rho^{T-1} + \text{coefficient } \times \rho^{T-3} + \ldots\ .$$

If T is odd, the first factor in the second term is

$$(3.27)\qquad \sum_{s=1}^{\frac{T-1}{2}}\left[d_s(1-\rho d_s) - d_s(1+\rho d_s)\right]\prod_{\substack{q=1\\ q\neq s}}^{\frac{T-1}{2}}(1-\rho^2 d_q^2)$$

$$= -2\sum_{s=1}^{\frac{T-1}{2}}\rho d_s^2\prod_{\substack{q=1\\ q\neq s}}^{\frac{T-1}{2}}(1-\rho^2 d_q^2)$$

$$= (T-1)\prod_{s=1}^{\frac{T-1}{2}}(-d_s^2)\rho^{T-2} + \text{coefficient } \times \rho^{T-4} + \dots$$
$$= (-1)^{\frac{T-1}{2}}(T-1)\frac{T+1}{2}\ \rho^{T-2} + \text{coefficient } \times \rho^{T-4} + \dots$$
$$= (-1)^{\frac{T-1}{2}}\frac{T^2-1}{2}\ \rho^{T-2} + \text{coefficient } \times \rho^{T-4} + \dots .$$

We use the fact that $\prod_{t\neq\frac{T+1}{2}} d_t$ is the coefficient of ρ^{T-1} in $|R|$.

If T is even, the second factor is

$$(3.28)\quad \sum_{t=1}^{T} z_t^2 \prod_{r\neq t}(1+\rho d_r) = \sum_{t=1}^{\frac{T}{2}}\left[z_t^2(1-\rho d_t) + z_{T+1-t}^2(1+\rho d_t)\right]\prod_{\substack{r=1\\ r\neq t}}^{\frac{T}{2}}(1-\rho^2 d_r^2)$$
$$= \sum_{t=1}^{\frac{T}{2}}\left[z_t^2 + z_{T+1-t}^2 - \rho d_t(z_t^2 - z_{T+1-t}^2)\right]\prod_{\substack{r=1\\ r\neq t}}^{\frac{T}{2}}(1-\rho^2 d_r^2)$$
$$= -\sum_{t=1}^{\frac{T}{2}}(z_t^2 - z_{T+1-t}^2)d_t \prod_{\substack{r=1\\ r\neq t}}^{\frac{T}{2}}(-d_r^2)\rho^{T-1}$$
$$+ \sum_{t=1}^{\frac{T}{2}}(z_t^2 + z_{T+1-t}^2)\prod_{\substack{r=1\\ r\neq t}}^{\frac{T}{2}}(-d_r^2)\rho^{T-2} + \dots$$
$$= \sum_{t=1}^{\frac{T}{2}}\frac{z_t^2 - z_{T+1-t}^2}{d_t}\ |G|\ \rho^{T-1}$$
$$- \sum_{t=1}^{\frac{T}{2}}\frac{z_t^2 + z_{T+1-t}^2}{d_t^2}\ |G|\ \rho^{T-2} + \dots$$
$$= (-1)^{\frac{T}{2}}\sum_{t=1}^{T}\frac{z_t^2}{d_t}\ \rho^{T-1} - (-1)^{\frac{T}{2}}\sum_{t=1}^{T}\frac{z_t^2}{d_t^2}\ \rho^{T-2} + \dots .$$

If T is odd, the second factor is

$$(3.29)\quad \sum_{t=1}^{\frac{T-1}{2}}\left[z_t^2 + z_{T+1-t}^2 - \rho d_t(z_t^2 - z_{T+1-t}^2)\right]\prod_{\substack{r=1\\ r\neq t}}^{\frac{T-1}{2}}(1-\rho^2 d_r^2)$$
$$+ z_{\frac{T+1}{2}}^2 \prod_{r=1}^{\frac{T-1}{2}}(1-\rho^2 d_r^2) = \prod_{r=1}^{\frac{T-1}{2}}(-d_r^2)z_{\frac{T+1}{2}}^2\ \rho^{T-1}$$
$$+ \dots = (-1)^{\frac{T-1}{2}}\frac{T+1}{2}z_{\frac{T+1}{2}}^2\ \rho^{T-1} + \dots .$$

For T even, the left-hand side of (3.23) is

$$-T\sum_{t=1}^{T}\frac{z_t^2}{d_t^2}\ \rho^{2T-3}+\dots\ . \tag{3.30}$$

For T odd, the left-hand side of (3.23) is

$$-\frac{(T-1)(T+1)^2}{4}z^2_{\frac{T+1}{2}}\ \rho^{2T-3}+\dots\ . \tag{3.31}$$

With probability 1 the derivative equation is of degree $2T-3$ for every T.

The coefficients of the polynomial in ρ in (3.23) are linear functions of $z_1^2,\dots,z_T^2$, which are independent χ_1^2 variables, hence the set of variables has a density. The roots being simple is the complement to some roots being multiple. The latter event is described by some algebraic relations among the coefficients and hence among $z_1^2,\dots,z_T^2$. The event has Lebesgue measure zero and hence probability zero. Hence, with probability 1 the degree of the polynomial equation (3.23) is odd, the number of real roots is odd, and they are distinct.

At each root the derivative is 0; hence, a local maximum or minimum occurs at the point; further, $n(\rho)\geq 0$ and $n(\rho)\to 0$ as $\rho\to\pm a$. Thus the number of maxima is one more than the number of minima.

The relative maxima of $n(\rho)$ occur for $\rho\in(-a,a)$. The relative maxima of

$$n^*(\alpha)=n\left(\frac{\alpha}{1+\alpha^2}\right) \tag{3.32}$$

occur for $\alpha\in[-1,1]$. Let $\rho_1<\dots<\rho_K$ be the values of ρ for which maxima occur. The probability that one of these values is $\pm\frac{1}{2}$ is 0 (by the above argument); those events can be ignored. If $-\frac{1}{2}<\rho_j<\frac{1}{2}$,

$$\alpha=\frac{1-\sqrt{1-4\rho_j^2}}{2\rho_j} \tag{3.33}$$

is real and yields a relative maximum of $n^*(\alpha)$. If $|\rho_j|>\frac{1}{2}$, the solution to (3.33) is not real and hence does not correspond to a maximum of $n^*(\alpha)$. Hence, the number of maxima with respect to α can be smaller than with respect to ρ. Anderson and Takemura (1986) showed that the root $\widehat{\alpha}=1$ yields a relative maximum of $n^*(\alpha)$ if $d\log n(\rho)/d\rho>0$ at $\rho=\frac{1}{2}$. (Alternatively, the root $\widehat{\alpha}=-1$ yields a relative maximum if the derivative is negative at $\rho=-\frac{1}{2}$).

Let $\eta_1 < \cdots < \eta_{K-1}$ be the values of ρ for which minima of $n(\rho)$ occur. Then

$$\rho_1 < \eta_1 < \cdots < \eta_{K-1} < \rho_K. \tag{3.34}$$

If $\eta_{k-1} < \frac{1}{2} < \rho_k$ for some $k (k = 2, \ldots, K)$, then $dn/d\rho > 0$ at $\rho = \frac{1}{2}$ and there is a relative maximum of $n^*(\alpha)$ at $\alpha = 1$. If $\rho_{k-1} < \frac{1}{2} < \eta_k$ then $dn/d\rho < 0$ at $\rho = \frac{1}{2}$ and $\alpha = 1$ gives a minimum.

Let $\alpha_1 < \cdots < \alpha_J$ be the values of α giving relative maxima of $n^*(\alpha)$. Each satisfies (3.33) except possibly $\alpha_1 = -1$ and/or $\alpha_J = 1$. The last may occur only if $\rho_K > \frac{1}{2}$. Thus $J \le K$. The maximum likelihood estimate of α is that one of $\alpha_1, \ldots, \alpha_J$ for which $n^*(\alpha_j)$ is greatest.

The maximum likelihood estimate of ρ is that one of $\rho_1, \ldots, \rho_K$ for which $n(\rho_j)$ is greatest. If that ρ_j is in $(-\frac{1}{2}, \frac{1}{2})$, then the maximum likelihood estimate of α is given by (3.33) for that ρ_j. If the maximizing ρ_j is outside $(-\frac{1}{2}, \frac{1}{2})$ (and hence $\rho_1 < -\frac{1}{2}$ or $\rho_K > \frac{1}{2}$), the maximizing α may be a solution to (3.33) for another ρ_j or it may be -1 or 1. If $n(\frac{1}{2}) > n(\rho_j)$ for every $\rho_j \in (-\frac{1}{2}, \frac{1}{2})$ and $n(\frac{1}{2}) > n(-\frac{1}{2})$ the maximizing α is $\alpha = 1$; if $n(-\frac{1}{2}) > n(\rho_j)$ for every $\rho_j \in (-\frac{1}{2}, \frac{1}{2})$ and $n(-\frac{1}{2}) > n(\frac{1}{2})$, the maximizing α is $\alpha = -1$. If $n(\frac{1}{2}) < n(\rho_j)$ and $n(-\frac{1}{2}) < n(\rho_j)$ for some j, then the maximizing α is (3.33) for some j.

Anderson and Takemura (1986) evaluated the probability that $\alpha = 1$ and alternatively $\alpha = -1$ yield relative maxima. The probability that $\widehat{\alpha} = 1$ or $\widehat{\alpha} = -1$ is, of course, less than the probability that a relative maximum occurs at $\alpha = 1$ or $\alpha = -1$, respectively.

4. Iterative Procedures Derived from Newton-Raphson and Scoring Methods

In this section we derive several iterative procedures to estimate the parameters of model (2.1), with emphasis on the estimation of ρ. After some preliminaries and the introduction of some notation, we operate successively in the time and frequency domains.

4.1. Notation and general rules

To estimate by maximum likelihood the covariances of the moving average part and the coefficients of the autoregressive part of an ARMA(p,q) model, Anderson (1977), Section 4.1, derived the equations that in general correspond to the iterative procedures in the Gaussian case. In the case of a MA(q) model these equations are

$$\widehat{\boldsymbol{A}}_{i-1}(\widehat{\boldsymbol{\sigma}}_i - \widehat{\boldsymbol{\sigma}}_{i-1}) = \widehat{\boldsymbol{s}}_{i-1}, \tag{4.1}$$

where $\widehat{\boldsymbol{\sigma}}_i = (\widehat{\sigma}_0^{(i)}, \widehat{\sigma}_1^{(i)}, \dots, \widehat{\sigma}_q^{(i)})'$ is the vector of estimated covariances at step i, $\widehat{\boldsymbol{\Lambda}}_{i-1}$ is an estimate of the information matrix for the covariances, and $\widehat{\boldsymbol{s}}_{i-1}$ is composed of estimates of $\partial \log L/\partial\sigma_j$, $j = 0, 1, \dots, q$. From (4.1) we deduce the iterative procedure

$$\widehat{\boldsymbol{\Lambda}}_{i-1}\widehat{\boldsymbol{\sigma}}_i = \widehat{\boldsymbol{s}}_{i-1} + \widehat{\boldsymbol{\Lambda}}_{i-1}\widehat{\boldsymbol{\sigma}}_{i-1} \equiv \widehat{\boldsymbol{r}}_{i-1}, \tag{4.2}$$

that for the MA(1) model is written in terms of components as follows:

$$\begin{pmatrix} \widehat{\lambda}_{00}^{(i-1)} & \widehat{\lambda}_{01}^{(i-1)} \\ \widehat{\lambda}_{10}^{(i-1)} & \widehat{\lambda}_{11}^{(i-1)} \end{pmatrix} \begin{pmatrix} \widehat{\sigma}_0^{(i)} \\ \widehat{\sigma}_1^{(i)} \end{pmatrix} = \begin{pmatrix} \widehat{r}_0^{(i-1)} \\ \widehat{r}_1^{(i-1)} \end{pmatrix}. \tag{4.3}$$

Solving for $\widehat{\sigma}_0^{(i)}$ and $\widehat{\sigma}_1^{(i)}$ and using $\widehat{\rho}^{(i)} = \widehat{\sigma}_1^{(i)}/\widehat{\sigma}_0^{(i)}$, we obtain the iterative procedure

$$\left\{\widehat{\lambda}_{01}^{(i-1)}\widehat{r}_1^{(i-1)} - \widehat{\lambda}_{11}^{(i-1)}\widehat{r}_0^{(i-1)}\right\}\widehat{\rho}^{(i)} = \widehat{\lambda}_{10}^{(i-1)}\widehat{r}_0^{(i-1)} - \widehat{\lambda}_{00}^{(i-1)}\widehat{r}_1^{(i-1)}. \tag{4.4}$$

The needed λ_{ij} and r_i can be evaluated by using the scoring or Newton-Raphson procedures, as will be shown next.

The derivation of these procedures will lead to certain quadratic forms and traces that we now consider. We use the fact that if $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices with $\boldsymbol{B}$ nonsingular, then $\boldsymbol{AB} = \boldsymbol{BA}$ implies that $\boldsymbol{AB}^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}$. We use this result with $\boldsymbol{A} = \boldsymbol{G}$ and $\boldsymbol{B} = \boldsymbol{R}$, for example. From (2.9) we use $\boldsymbol{R} = \boldsymbol{I} + \rho\boldsymbol{G}$, and hence,

$$\boldsymbol{RG} = (\boldsymbol{I} + \rho\boldsymbol{G})\boldsymbol{G} = \boldsymbol{G} + \rho\boldsymbol{G}^2 = \boldsymbol{G}(\boldsymbol{I} + \rho\boldsymbol{G}) = \boldsymbol{GR}. \tag{4.5}$$

We now define a set of quadratic forms by

$$q_{jk} = \boldsymbol{y}'\boldsymbol{R}^{-(j+1)}\boldsymbol{G}^k\boldsymbol{y}, \quad j = -1, 0, 1, \dots; \quad k = 0, 1, 2, \dots\ , \tag{4.6}$$

and a set of traces by

$$t_{jk} = \text{tr}\ \boldsymbol{R}^{-j}\boldsymbol{G}^k, \quad j, k = 0, 1, 2, \dots\ . \tag{4.7}$$

We note that q_{jk} is a random variable, a function of $\boldsymbol{y}$, and that

$$\begin{aligned} \mathcal{E}q_{jk} &= \mathcal{E}\boldsymbol{y}'\boldsymbol{R}^{-(j+1)}\boldsymbol{G}^k\boldsymbol{y} = \mathcal{E}\ \text{tr}\ \boldsymbol{y}'\boldsymbol{R}^{-(j+1)}\boldsymbol{G}^k\boldsymbol{y} = \text{tr}\ \boldsymbol{R}^{-(j+1)}\boldsymbol{G}^k\mathcal{E}\boldsymbol{y}\boldsymbol{y}' \\ &= \text{tr}\ \boldsymbol{R}^{-(j+1)}\boldsymbol{G}^k\boldsymbol{\Sigma} = \text{tr}\ \boldsymbol{R}^{-(j+1)}\boldsymbol{G}^k\sigma_0\boldsymbol{R} = \sigma_0\ \text{tr}\ \boldsymbol{R}^{-j}\boldsymbol{G}^k \\ &= \sigma_0 t_{jk}, \end{aligned} \tag{4.8}$$

where we used (2.6) and the fact that $\boldsymbol{G}$ and $\boldsymbol{R}$ commute.

In our operations we shall find quadratic forms like $\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y}$, $\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{y}$, etc., and traces like tr $\boldsymbol{R}^{-2}\boldsymbol{G}$, tr $\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}$, etc., so that frequent use of this commutative property will be made.

4.2. Time domain considerations

4.2.1. Procedures based on the likelihood function

Iterative Procedure 1 (Time, Likelihood Function, Scoring)

In Anderson (1977) Section 4.2, it is shown that (4.1) above leads to a system that in the case of a MA(1) model is

$$
\begin{aligned}
&\operatorname{tr}\, \widehat{\boldsymbol{\Sigma}}_{i-1}^{-2} \widehat{\sigma}_0^{(i)} + \operatorname{tr}\, \widehat{\boldsymbol{\Sigma}}_{i-1}^{-2} \boldsymbol{G} \widehat{\sigma}_1^{(i)} = \boldsymbol{y}' \widehat{\boldsymbol{\Sigma}}_{i-1}^{-2} \boldsymbol{y}, \\
&\operatorname{tr}\, \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \widehat{\sigma}_0^{(i)} + \operatorname{tr}\, \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \widehat{\sigma}_1^{(i)} = \boldsymbol{y}' \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{y}.
\end{aligned} \tag{4.9}
$$

These equations can also be written with $\boldsymbol{R}$ instead of $\boldsymbol{\Sigma}$, using the fact that $\boldsymbol{\Sigma} = \sigma_0 \boldsymbol{R}$, and σ_0^{-2} cancelled throughout. We can then write the resulting linear system as

$$
\begin{pmatrix} \widehat{t}_{20}^{(i-1)} & \widehat{t}_{21}^{(i-1)} \\ \widehat{t}_{21}^{(i-1)} & \widehat{t}_{22}^{(i-1)} \end{pmatrix} \begin{pmatrix} \widehat{\sigma}_0^{(i)} \\ \sigma_1^{(i)} \end{pmatrix} = \begin{pmatrix} \widehat{q}_{10}^{(i-1)} \\ \widehat{q}_{11}^{(i-1)} \end{pmatrix}. \tag{4.10}
$$

Solving this system for $\widehat{\sigma}_0^{(i)}$ and $\widehat{\sigma}_1^{(i)}$, and using the definition of $\widehat{\rho}^{(i)}$, we obtain the following iterative procedure:

$$
\left\{ \widehat{t}_{22}^{(i-1)} \widehat{q}_{10}^{(i-1)} - \widehat{t}_{21}^{(i-1)} \widehat{q}_{11}^{(i-1)} \right\} \widehat{\rho}^{(i)} = \widehat{t}_{20}^{(i-1)} \widehat{q}_{11}^{(i-1)} - \widehat{t}_{21}^{(i-1)} \widehat{q}_{10}^{(i-1)}. \tag{4.11}
$$

Iterative Procedure 2 (Time, Likelihood Function, Newton-Raphson)

Equations (4.2) in Anderson (1975) become, in the case of a MA(1) model,

$$
\begin{aligned}
&\left(\boldsymbol{y}' \widehat{\boldsymbol{\Sigma}}_{i-1}^{-3} \boldsymbol{y} - \frac{1}{2} \operatorname{tr}\, \widehat{\boldsymbol{\Sigma}}_{i-1}^{-2} \right) \widehat{\sigma}_0^{(i)} \\
&\qquad + \left(\boldsymbol{y}' \widehat{\boldsymbol{\Sigma}}_{i-1}^{-2} \boldsymbol{G} \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{y} - \frac{1}{2} \operatorname{tr}\, \widehat{\boldsymbol{\Sigma}}_{i-1}^{-2} \boldsymbol{G} \right) \widehat{\sigma}_1^{(i)} = \frac{3}{2} \boldsymbol{y}' \widehat{\boldsymbol{\Sigma}}_{i-1}^{-2} \boldsymbol{y} - \operatorname{tr}\, \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1}, \\
&\left(\boldsymbol{y}' \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \widehat{\boldsymbol{\Sigma}}_{i-1}^{-2} \boldsymbol{y} - \frac{1}{2} \operatorname{tr}\, \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \right) \widehat{\sigma}_0^{(i)} \\
&\qquad - \left(\boldsymbol{y}' \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{y} - \frac{1}{2} \operatorname{tr}\, \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \right) \widehat{\sigma}_1^{(i)} \\
&\qquad\qquad = \frac{3}{2} \boldsymbol{y}' \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G} \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{y} - \operatorname{tr}\, \widehat{\boldsymbol{\Sigma}}_{i-1}^{-1} \boldsymbol{G}.
\end{aligned} \tag{4.12}
$$

Since $\boldsymbol{G}$ commutes with $\boldsymbol{\Sigma}$ ($\boldsymbol{\Sigma} = \sigma_0 \boldsymbol{R}$), it also commutes with $\boldsymbol{\Sigma}^{-1}$. Introducing the notation

$$
q_{jk}^* = \boldsymbol{y}' \boldsymbol{\Sigma}^{-(j+1)} \boldsymbol{G}^k \boldsymbol{y}, \qquad t_{jk}^* = \operatorname{tr}\, \boldsymbol{\Sigma}^{-j} \boldsymbol{G}^k, \tag{4.13}
$$

so that $\mathcal{E}q^*_{jk} = t^*_{jk}$, we write (4.12) in matrix form as

$$(4.14)\qquad \begin{pmatrix} \widehat{q}^{*(i-1)}_{20} - \frac{1}{2}\widehat{t}^{*(i-1)}_{20} & \widehat{q}^{*(i-1)}_{21} - \frac{1}{2}\widehat{t}^{*(i-1)}_{21} \\ \widehat{q}^{*(i-1)}_{21} - \frac{1}{2}\widehat{t}^{*(i-1)}_{21} & \widehat{q}^{*(i-1)}_{22} - \frac{1}{2}\widehat{t}^{*(i-1)}_{22} \end{pmatrix} \begin{pmatrix} \widehat{\sigma}^{(i)}_0 \\ \widehat{\sigma}^{(i)}_1 \end{pmatrix} = \begin{pmatrix} \frac{3}{2}\widehat{q}^{*(i-1)}_{10} - \widehat{t}^{*(i-1)}_{10} \\ \frac{3}{2}\widehat{q}^{*(i-1)}_{11} - \widehat{t}^{*(i-1)}_{11} \end{pmatrix}.$$

Since $\widehat{\boldsymbol{\Sigma}}^{-1}$ appears in the coefficients of the system (4.12) raised to different powers, substitution of $\boldsymbol{\Sigma} = \sigma_0 \boldsymbol{R}$ will not produce the cancellation of all powers of σ_0^{-1}. Hence, we do not write an expression for ρ as we did in Iterative Procedure 1, but leave the iterations to be carried out for σ_0 and σ_1 as indicated by (4.14).

4.2.2. Procedures based on the concentrated likelihood function

Iterative Procedure 3 (Time, Concentrated Likelihood Function, Scoring)

We operate with (3.10), that is,

$$(4.15)\qquad \log n(\rho) = -\log|\boldsymbol{I} + \rho\boldsymbol{G}| - T\log\{\boldsymbol{y}'(\boldsymbol{I} + \rho\boldsymbol{G})^{-1}\boldsymbol{y}\}.$$

We have

$$(4.16)\qquad \begin{aligned} \frac{d}{d\rho}\log n(\rho) &= -\,\mathrm{tr}\ \boldsymbol{R}^{-1}\boldsymbol{G} + T\frac{\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}} \\ &= -\frac{(\mathrm{tr}\ \boldsymbol{R}^{-1}\boldsymbol{G})(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}) - T\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}} \\ &= 0. \end{aligned}$$

To apply the scoring method we use the fact that

$$(4.17)\qquad \boldsymbol{I} = \boldsymbol{R}^{-1}\boldsymbol{R} = \boldsymbol{R}^{-1}(\boldsymbol{I} + \rho\boldsymbol{G}) = \boldsymbol{R}^{-1} + \rho\boldsymbol{R}^{-1}\boldsymbol{G},$$

so that

$$(4.18)\qquad \mathrm{tr}\ \boldsymbol{R}^{-1}\boldsymbol{G} = \mathrm{tr}\ \boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{R}^{-1} + \rho\boldsymbol{R}^{-1}\boldsymbol{G}) = \mathrm{tr}\ \boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1} + \rho\ \mathrm{tr}\ \boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{G}.$$

Substitution in the numerator of (4.16) gives the estimating equation

$$(4.19)\qquad \begin{aligned} &\mathrm{tr}\ \widehat{\boldsymbol{R}}^{-1}_{i-1}\boldsymbol{G}\widehat{\boldsymbol{R}}^{-1}_{i-1}\boldsymbol{G}(\widehat{\boldsymbol{y}'\widehat{\boldsymbol{R}}^{-1}_{i-1}\boldsymbol{y}})\widehat{\rho}^{(i)} \\ &\qquad = T\boldsymbol{y}'\widehat{\boldsymbol{R}}^{-1}_{i-1}\boldsymbol{G}\widehat{\boldsymbol{R}}^{-1}_{i-1}\boldsymbol{y} - \mathrm{tr}\ \widehat{\boldsymbol{R}}^{-1}_{i-1}\boldsymbol{G}\widehat{\boldsymbol{R}}^{-1}(\boldsymbol{y}'\widehat{\boldsymbol{R}}^{-1}_{i-1}\boldsymbol{y}), \end{aligned}$$

or, in the notation introduced in (4.6) and (4.7),

$$\widehat{t}_{22}^{(i-1)}\widehat{q}_{00}^{(i-1)}\widehat{\rho}^{(i)} = T\widehat{q}_{11}^{(i-1)} - \widehat{t}_{21}^{(i-1)}\widehat{q}_{00}^{(i-1)}. \tag{4.20}$$

Iterative Procedure 4 (Time, Concentrated Likelihood Function, Scoring)

To derive Iterative Procedure 3 we used (4.17). Another approach is to go back to the general structure given by (3.12). The first derivative of log $n(\rho)$ with respect to ρ is given by (4.16), and the second derivative is

$$\begin{aligned}(4.21)\quad &\frac{d^2}{d\rho^2}\log n(\rho)\\ &= \operatorname{tr} \boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{G} + T\frac{-2\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y}(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}) + (\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y})^2}{(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y})^2}\\ &= \frac{\operatorname{tr} \boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y})^2 - 2T\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y}(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}) + T(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y})^2}{(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y})^2}.\end{aligned}$$

The method of scoring consists in replacing (4.21) by its expected value; instead, we can replace each quadratic form in (4.21) by its expected value, using (4.8). In the notation of (4.6) and (4.17) this is

$$t_{22} + T\frac{-2\sigma_0 t_{22}\sigma_0 T + (\sigma_0 t_{11})^2}{(\sigma_0 T)^2} = -t_{22} + \frac{t_{11}^2}{T}. \tag{4.22}$$

Finally, Iterative Procedure 4 to estimate ρ is given by

$$\begin{aligned}(4.23)\quad \widehat{q}_{00}^{(i-1)}\left\{\widehat{t}_{22}^{(i-1)} - \frac{1}{T}\left[\widehat{t}_{11}{}^{(i-1)}\right]^2\right\}\widehat{\rho}^{(i)} &= T\widehat{q}_{11}^{(i-1)} - \widehat{t}_{11}^{(i-1)}\widehat{q}_{00}^{(i-1)}\\ &\quad + \widehat{q}_{00}^{(i-1)}\left\{\widehat{t}_{22}^{(i-1)} - \frac{1}{T}\left[\widehat{t}_{11}^{(i-1)}\right]^2\right\}\widehat{\rho}^{(i-1)}.\end{aligned}$$

Iterative Procedure 5 (Time, Concentrated Likelihood Function, Newton-Raphson)

The iterative procedure is (3.12). From (4.16)

$$\begin{aligned}(4.24)\quad (\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y})^2\frac{d}{d\rho}\log n(\rho) &= T(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y})(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}) - \operatorname{tr} \boldsymbol{R}^{-1}\boldsymbol{G}(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y})^2\\ &= Tq_{11}q_{00} - t_{11}q_{00}^2.\end{aligned}$$

From (4.21)

$$-(\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y})^2 \frac{d^2}{d\rho^2} \log n(\rho) = 2Tq_{22}q_{00} - t_{22}q_{00}^2 - Tq_{11}^2. \tag{4.25}$$

It follows that the iterative procedure for ρ can be written as

$$\begin{aligned}\left\{2T\hat{q}_{22}^{(i-1)}\hat{q}_{00}^{(i-1)} - \hat{t}_{22}^{(i-1)}\left[\hat{q}_{00}^{(i-1)}\right]^2 - T\left[\hat{q}_{11}^{(i-1)}\right]^2\right\}\hat{\rho}^{(i)} \\ = T\hat{q}_{11}^{(i-1)}\hat{q}_{00}^{(i-1)} - \hat{t}_{11}^{(i-1)}\left[\hat{q}_{00}^{(i-1)}\right]^2 \\ + \left\{2T\hat{q}_{22}^{(i-1)}\hat{q}_{00}^{(i-1)} - \hat{t}_{22}^{(i-1)}\left[\hat{q}_{00}^{(i-1)}\right]^2 - T\left[\hat{q}_{11}^{(i-1)}\right]^2\right\}\hat{\rho}^{(i-1)}.\end{aligned} \tag{4.26}$$

4.3. Frequency domain considerations

In this section we write all quadratic forms and traces appearing in the iterative procedures presented in Section 4.2., in terms of the elements introduced in paragraph III of Section 3.

From (3.7) we have $\boldsymbol{I} = \boldsymbol{K}\boldsymbol{K}'$ and $\boldsymbol{G} = \boldsymbol{K}\boldsymbol{D}\boldsymbol{K}'$, where $\boldsymbol{D}$ is diagonal with diagonal elements $d_j = 2\cos[\pi j/(T+1)]$, $j = 1, 2, \ldots, T$. Then

$$\boldsymbol{R} = \boldsymbol{I} + \rho\boldsymbol{G} = \boldsymbol{K}\boldsymbol{K}' + \rho\boldsymbol{K}\boldsymbol{D}\boldsymbol{K}' = \boldsymbol{K}(\boldsymbol{I} + \rho\boldsymbol{D})\boldsymbol{K}', \tag{4.27}$$

so that $\boldsymbol{R}$ is also diagonalized by the orthogonal matrix $\boldsymbol{K}$, and $\boldsymbol{I} + \rho\boldsymbol{D}$ has diagonal elements $1 + 2\rho\cos[\pi j/(T+1)]$. It then follows that

$$\boldsymbol{R}^{-s} = \boldsymbol{K}(\boldsymbol{I} + \rho\boldsymbol{D})^{-s}\boldsymbol{K}', \qquad \boldsymbol{G}^s = \boldsymbol{K}\boldsymbol{D}^s\boldsymbol{K}', \qquad s = 0, 1, \ldots\ . \tag{4.28}$$

Putting these results together we find that the quadratic forms introduced in (4.6) can be written in terms of the z_j defined in (3.8) as

$$\begin{aligned} q_{jk} &= \boldsymbol{y}'\boldsymbol{R}^{-(j+1)}\boldsymbol{G}^k\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{K}(\boldsymbol{I} + \rho\boldsymbol{D})^{-(j+1)}\boldsymbol{K}'\boldsymbol{K}\boldsymbol{D}^k\boldsymbol{K}'\boldsymbol{y} \\ &= (\boldsymbol{K}'\boldsymbol{y})'(\boldsymbol{I} + \rho\boldsymbol{D})^{-(j+1)}\boldsymbol{D}^k(\boldsymbol{K}'\boldsymbol{y}) = \boldsymbol{z}'(\boldsymbol{I} + \rho\boldsymbol{D})^{-(j+1)}\boldsymbol{D}^k\boldsymbol{z} \\ &= \sum_{s=1}^{T} \frac{d_s^k}{(1 + \rho d_s)^{j+1}} z_s^2, \qquad j = -1, 0, 1, \ldots;\ k = 0, 1, \ldots\ , \end{aligned} \tag{4.29}$$

while the traces introduced in (4.7) become

$$\begin{aligned} t_{jk} &= \operatorname{tr} \boldsymbol{R}^{-j}\boldsymbol{G}^k = \operatorname{tr} \boldsymbol{K}(\boldsymbol{I} + \rho\boldsymbol{D})^{-j}\boldsymbol{G}^k\boldsymbol{K}' \\ &= \operatorname{tr}(\boldsymbol{I} + \rho\boldsymbol{D})^{-j}\boldsymbol{D}^k\boldsymbol{K}'\boldsymbol{K} = \operatorname{tr}(\boldsymbol{I} + \rho\boldsymbol{D})^{-j}\boldsymbol{D}^k \\ &= \sum_{s=1}^{T} \frac{d_s^k}{(1 + \rho d_s)^j}, \qquad j, k = 0, 1, \ldots\ . \end{aligned} \tag{4.30}$$

For an interesting connection of these results with the analysis of variance see the paper by Speed (1987) and the comments by Anderson (1987).

5. Evaluation of Quadratic Forms in the Time Domain

5.1. Introduction

In this section, we develop some algebraic procedures for the evaluation of quadratic forms. In Section 9 we compare some of these from the point of view of efficient computation.

We first show that all quadratic forms q_{jk} used in Section 4, where $j \geq k$, can be expressed as functions of $q_{j0} = \boldsymbol{y}'\boldsymbol{R}^{-(j+1)}\boldsymbol{y}$. In effect, since $\boldsymbol{R} = \boldsymbol{I} + \rho\boldsymbol{G}$, we can substitute $\boldsymbol{G} = \rho^{-1}(\boldsymbol{R} - \boldsymbol{I})$ in q_{jk}, provided $\rho \neq 0$, to obtain

$$
\begin{aligned}
(5.1) \qquad q_{jk} &= \boldsymbol{y}'\boldsymbol{R}^{-(j+1)}\boldsymbol{G}^k\boldsymbol{y} = \rho^{-k}\boldsymbol{y}'\boldsymbol{R}^{-(j+1)}(\boldsymbol{R} - \boldsymbol{I})^k\boldsymbol{y} \\
&= \rho^{-k}\boldsymbol{y}'\boldsymbol{R}^{-(j+1)}\left\{\sum_{s=0}^{k}(-1)^{k-s}\binom{k}{s}\boldsymbol{R}^s\right\}\boldsymbol{y} \\
&= \rho^{-k}\sum_{s=0}^{k}(-1)^{k-s}\binom{k}{s}\boldsymbol{y}'\boldsymbol{R}^{-(j-s+1)}\boldsymbol{y} \\
&= \rho^{-k}\sum_{s=0}^{k}(-1)^{k-s}\binom{k}{s}q_{j-s,0}, \qquad j \geq k.
\end{aligned}
$$

For example,

$$
(5.2) \qquad q_{11} = \frac{1}{\rho}(q_{00} - q_{10}), \quad q_{21} = \frac{1}{\rho}(q_{10} - q_{20}), \quad q_{22} = \frac{1}{\rho^2}(q_{00} - 2q_{10} + q_{20}).
$$

These relations can be used to express the iterative procedures of Section 4 as functions of the various traces and of the q_{j0}. For example, in Iterative Procedure 1, (4.11) becomes

$$
\begin{aligned}
(5.3) \qquad &\left\{\left[\hat{\rho}^{(i-1)}\hat{t}_{22}^{(i-1)} + \hat{t}_{21}^{(i-1)}\right]\hat{q}_{10}^{(i-1)} - \hat{t}_{21}^{(i-1)}\hat{q}_{00}^{(i-1)}\right\}\hat{\rho}^{(i)} \\
&\qquad = \hat{t}_{20}^{(i-1)}\hat{q}_{00}^{(i-1)} - \left[\hat{\rho}^{(i-1)}\hat{t}_{21}^{(i-1)} + \hat{t}_{20}^{(i-1)}\right]\hat{q}_{10}^{(i-1)},
\end{aligned}
$$

while in Iterative Procedure 3, (4.20) becomes

$$
(5.4) \qquad \hat{\rho}^{(i-1)}\hat{t}_{22}^{(i-1)}\hat{q}_{00}^{(i-1)}\hat{\rho}^{(i)} = \left[T - \hat{\rho}^{(i-1)}\hat{t}_{21}^{(i-1)}\right]\hat{q}_{00}^{(i-1)} - T\hat{q}_{10}^{(i-1)}.
$$

Let us now define

$$x = R^{-1}y, \qquad v = R^{-1}x. \tag{5.5}$$

Then

$$q_{00} = y'R^{-1}y = y'x, \quad q_{10} = y'R^{-2}y = x'x, \quad q_{20} = y'R^{-3}y = y'v, \tag{5.6}$$

and we see that it suffices to solve for x in the linear system

$$y = Rx, \tag{5.7}$$

and having done that to solve for v the linear system

$$x = Rv. \tag{5.8}$$

Once y, x, and v are available, all quadratic forms appearing in the iterative procedures defined in Section 4 can be easily expressed in terms of the components of these vectors. In effect,

$$q_{00} = \sum_{i=1}^{T} y_i x_i, \quad q_{10} = \sum_{i=1}^{T} x_i^2, \quad q_{20} = \sum_{i=1}^{T} x_i v_i, \tag{5.9}$$

while

$$q_{11} = y'R^{-2}Gy = y'R^{-1}GR^{-1}y = x'Gx = 2\sum_{i=1}^{T-1} x_i x_{i+1}, \tag{5.10}$$

$$q_{21} = y'R^{-3}Gy = x'R^{-1}Gx = v'Gx = \sum_{i=1}^{T-1}(x_{i+1}v_i + x_i v_{i+1}), \tag{5.11}$$

$$q_{22} = y'R^{-3}G^2y = v'G^2x = x_1v_1 + x_Tv_T + 2\sum_{i=2}^{T-1} x_i v_i + \sum_{i=1}^{T-2}(x_i v_{i+2} + x_{i+2}v_i). \tag{5.12}$$

Hence, it folows that the calculation of the quadratic forms can be reduced to the calculation of the q_{j0} for $j = 0, 1$ and 2; this in turn corresponds to solving explicitly for x the systems (5.7) and solving for v the system (5.8). This will be considered in the remaining parts of this section.

Useful references for the treatment of linear systems in the indicated context are Anderson (1984), in particular Appendix A, Golub and Van Loan (1983), and Graybill (1969).

5.2. Cholesky decomposition

The Cholesky decomposition of a matrix is a useful and efficient device, that in our case can help to compute the needed quadratic forms. We consider the decomposition of $\boldsymbol{R}$.

5.2.1. Derivations

Since $\boldsymbol{R}$ is symmetric and positive definite for $-a < \rho < a$, where $1/2 < a < 1$, in this range its unique Cholesky decomposition exists. It can be written as

$$\boldsymbol{R} = \boldsymbol{T}\boldsymbol{T}', \tag{5.13}$$

where $\boldsymbol{T} = (t_{ij})$ is bidiagonal (because $\boldsymbol{R}$ is tridiagonal), lower triangular, $t_{ii} > 0$, $t_{ij} = 0$ for $i < j$ and $i > j+1$. The decomposition can also be written as

$$\boldsymbol{R} = \boldsymbol{U}\boldsymbol{V}\boldsymbol{U}', \tag{5.14}$$

where $\boldsymbol{U} = (u_{ij})$ is bidiagonal, lower triangular, $u_{ii} = 1$, $u_{ij} = 0$ for $i < j$ and $i > j+1$, and $\boldsymbol{V} = (v_{ij})$ is diagonal with $v_{ii} > 0$ ($\boldsymbol{T} = \boldsymbol{U}\boldsymbol{V}^{\frac{1}{2}}$.)

Expression (5.13) is often called the Cholesky decomposition of $\boldsymbol{R}$, and $\boldsymbol{T}$ is called the Cholesky triangle. The procedure to obtain $\boldsymbol{T}$ is sometimes called the square root method. Setting $\boldsymbol{V}\boldsymbol{U}' = \boldsymbol{S}$, say, we see that $\boldsymbol{S}$ is bidiagonal, upper triangular and that $\boldsymbol{R} = \boldsymbol{U}\boldsymbol{S}$, which is a case of the so-called LR decomposistion.

Proposition 1. *The components of* $\boldsymbol{U}$ *and* $\boldsymbol{V}$ *in (5.14) satisfy*

$$v_{ss} = \frac{\Delta_s}{\Delta_{s-1}}, \qquad s = 1, \ldots, T, \tag{5.15}$$

and

$$u_{s+1,s} = \frac{\rho}{v_{ss}} = \rho\frac{\Delta_{s-1}}{\Delta_s}, \qquad s = 1, \ldots, T-1, \tag{5.16}$$

where the Δ_s *are given in (3.2).*

Proof. Let $\boldsymbol{R} = (r_{ij})$. In terms of components (5.14) is

$$r_{ij} = \sum_{s=1}^{T}\sum_{t=1}^{T} u_{is}v_{st}u_{jt} = \sum_{s=1}^{T} u_{is}v_{ss}u_{js} = u_{ii}v_{ii}u_{ji} + u_{i,i-1}v_{i-1,i-1}u_{j,i-1}. \tag{5.17}$$

For $j = i$ we have

$$1 = r_{ii} = u_{ii}^2 v_{ii} + u_{i,i-1}^2 v_{i-1,i-1} = v_{ii} + u_{i,i-1}^2 v_{i-1,i-1}; \tag{5.18}$$

for $j = i - 1$ we have

$$\rho = r_{i,i-1} = u_{ii}v_{ii}u_{i-1,i} + u_{i,i-1}v_{i-1,i-1}u_{i-1,i-1} = u_{i,i-1}v_{i-1,i-1}. \tag{5.19}$$

From this last expression we deduce that

$$u_{i,i-1} = \frac{\rho}{v_{i-1,i-1}}, \qquad i = 2, \ldots, T. \tag{5.20}$$

Using this expression in (5.18) we have

$$v_{ii} = 1 - \frac{\rho^2}{v_{i-1,i-1}} = \frac{v_{i-1,i-1} - \rho^2}{v_{i-1,i-1}}. \tag{5.21}$$

Direct evaluation provides the values

$$v_{11} = 1, \quad v_{22} = 1 - \rho^2, \quad v_{33} = \frac{1 - 2\rho^2}{1 - \rho^2}, \quad v_{44} = \frac{1 - 3\rho^2 + \rho^4}{1 - 2\rho^2}, \tag{5.22}$$

in agreement with (3.2) and (5.15). We then complete the proof by induction:

$$v_{ii} = \frac{\frac{\Delta_{i-1}}{\Delta_{i-2}} - \rho^2}{\frac{\Delta_{i-1}}{\Delta_{i-2}}} = \frac{\Delta_{i-1} - \rho^2\Delta_{i-2}}{\Delta_{i-1}} = \frac{\Delta_i}{\Delta_{i-1}}, \tag{5.23}$$

because the determinants of the $\boldsymbol{R}$ matrices of various order satisfy

$$\Delta_s = \Delta_{s-1} - \rho^2\Delta_{s-2}; \tag{5.24}$$

see Shaman (1969). Since further $\boldsymbol{U}$ and $\boldsymbol{V}$ are unique, the proof is completed. ∎

Note that (5.15) holds for the Cholesky decomposistion of any positive definite matrix $\boldsymbol{R}$. As defined in (2.8), $\boldsymbol{P}$ is symmetric and positive definite for any value of α. Hence,

its unique Cholesky decomposition exists and can be written in ways similar to (5.13) and (5.14), namely,

$$P = \widetilde{T}\widetilde{T}' = \widetilde{U}\widetilde{V}\widetilde{U}', \tag{5.25}$$

where the $\widetilde{T}$, $\widetilde{U}$, and $\widetilde{V}$ matrices have the same structure as for R. An argument similar to that in Proposition 2, together with the fact that in terms of α the determinants of the P matrices satisfy the relation $\tilde{\Delta}_s = (1+\alpha^2)\tilde{\Delta}_{s-1} - \alpha^2\tilde{\Delta}_{s-2}$, lead to the expressions

$$\tilde{v}_{ss} = \frac{\tilde{\Delta}_s}{\tilde{\Delta}_{s-1}} = \frac{1-\alpha^{2(s+1)}}{1-\alpha^{2s}}, \qquad s = 1, \dots, T, \tag{5.26}$$

$$\tilde{u}_{s+1,s} = \frac{\alpha}{\widetilde{v}_{ss}} = \alpha\frac{1-\alpha^{2s}}{1-\alpha^{2(s+1)}}, \qquad s = 1, \dots, T-1. \tag{5.27}$$

These are simpler expressions in α compared to those involving the polynomials in ρ given in (3.2).

Proposition 2. *The components of T in (5.13) satisfy*

$$t_{ss} = \sqrt{\frac{\Delta_s}{\Delta_{s-1}}}, \qquad s = 1, \dots, T\ , \tag{5.28}$$

$$t_{s,s-1} = \rho\sqrt{\frac{\Delta_{s-2}}{\Delta_{s-1}}}, \qquad s = 2, \dots, T\ . \tag{5.29}$$

Proof. Comparing (5.13) and (5.14) we see that $T = UV^{1/2}$, from which (5.28) and (5.29) follow. ∎

5.2.2. Using the Cholesky decomposition to compute quadratic forms

We now use the results of the previous section to derive expressions for q_{j0}, $j = 0, 1, 2$.

$$\begin{aligned} q_{00} &= y'R^{-1}y = y'(UVU')^{-1}y = (U^{-1}y)'V^{-1}(U^{-1}y) \\ &= w'V^{-1}w = \sum_{s=1}^{T}\frac{w_s^2}{v_{ss}} = \sum_{s=1}^{T}\frac{\Delta_{s-1}}{\Delta_s}w_s^2, \end{aligned} \tag{5.30}$$

where $\boldsymbol{w} = \boldsymbol{U}^{-1}\boldsymbol{y}$. It suffices to find $\boldsymbol{w}$ in the linear system $\boldsymbol{y} = \boldsymbol{U}\boldsymbol{w}$. This system is

$$(5.31) \qquad \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ u_{21} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & u_{T,T-1} & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_T \end{pmatrix},$$

and hence it follows that $w_1 = y_1$,

$$(5.32) \qquad w_s = y_s - u_{s,s-1} w_{s-1} = y_s - \rho \frac{\Delta_{s-2}}{\Delta_{s-1}} w_{s-1}, \qquad s = 2, \dots, T.$$

Let us now define $\boldsymbol{w}^* = \boldsymbol{V}^{-1}\boldsymbol{w} = (w_1/v_{11}, \dots, w_T/v_{TT})'$. Then,

$$(5.33) \qquad \begin{aligned} q_{10} &= \boldsymbol{y}'\boldsymbol{R}^{-2}\boldsymbol{y} = \boldsymbol{y}'(\boldsymbol{U}\boldsymbol{V}\boldsymbol{U}')^{-1}(\boldsymbol{U}\boldsymbol{V}\boldsymbol{U}')^{-1}\boldsymbol{y} = \boldsymbol{w}'\boldsymbol{V}^{-1}\boldsymbol{U}^{-1}\boldsymbol{U}'^{-1}\boldsymbol{V}^{-1}\boldsymbol{w} \\ &= \boldsymbol{w}^{*\prime}\boldsymbol{U}^{-1}(\boldsymbol{U}')^{-1}\boldsymbol{w}^* = \boldsymbol{x}'\boldsymbol{x} = \sum_{s=1}^{T} x_s^2, \end{aligned}$$

where

$$(5.34) \qquad (\boldsymbol{U}')^{-1}\boldsymbol{w}^* = (\boldsymbol{U}')^{-1}\boldsymbol{V}^{-1}\boldsymbol{w} = (\boldsymbol{U}')^{-1}\boldsymbol{V}^{-1}\boldsymbol{U}^{-1}\boldsymbol{y} = (\boldsymbol{U}\boldsymbol{V}\boldsymbol{U}')^{-1}\boldsymbol{y} = \boldsymbol{R}^{-1}\boldsymbol{y} = \boldsymbol{x},$$

as defined in (5.5). Hence, it suffices to find $\boldsymbol{x}$ in the linear system $\boldsymbol{w}^* = \boldsymbol{U}'\boldsymbol{x}$. This system is

$$(5.35) \qquad \begin{pmatrix} w_1/v_{11} \\ w_2/v_{22} \\ \vdots \\ w_{T-1}/v_{T-1,T-1} \\ w_T/v_{TT} \end{pmatrix} = \begin{pmatrix} 1 & u_{21} & 0 & \dots & 0 & 0 \\ 0 & 1 & u_{32} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & u_{T,T-1} \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{T-1} \\ x_T \end{pmatrix},$$

and it provides the recursive relations

$$(5.36) \qquad \begin{aligned} x_T &= \frac{w_T}{v_{TT}} = \frac{\Delta_{T-1}}{\Delta_T} w_T; \\ x_{T-s} &= \frac{w_{T-s} - \rho x_{T-s+1}}{v_{T-s,T-s}} = \frac{\Delta_{T-s-1}}{\Delta_{T-s}}(w_{T-s} - \rho x_{T-s+1}), \qquad s = 1, \dots, T-1. \end{aligned}$$

Finally,

$$(5.37) \qquad \begin{aligned} q_{20} &= \boldsymbol{y}'\boldsymbol{R}^{-3}\boldsymbol{y} = \boldsymbol{x}'(\boldsymbol{U}\boldsymbol{V}\boldsymbol{U}')^{-1}\boldsymbol{x} = (\boldsymbol{U}^{-1}\boldsymbol{x})'\boldsymbol{V}^{-1}(\boldsymbol{U}^{-1}\boldsymbol{x}) \\ &= \boldsymbol{h}'\boldsymbol{V}^{-1}\boldsymbol{h} = \sum_{s=1}^{T} \frac{h_s^2}{v_{ss}} = \sum_{s=1}^{T} \frac{\Delta_{s-1}}{\Delta_s} h_s^2, \end{aligned}$$

where $\boldsymbol{h} = \boldsymbol{U}^{-1}\boldsymbol{x}$. It suffices to solve for $\boldsymbol{h}$ the linear system $\boldsymbol{U}\boldsymbol{h} = \boldsymbol{x}$. This system is similar to (5.31) and hence it follows that $h_1 = x_1$,

$$h_s = x_s - u_{s,s-1}h_{s-1} = x_s - \rho\frac{\Delta_{s-2}}{\Delta_{s-1}}h_{s-1}, \qquad s = 2, \ldots, T. \tag{5.38}$$

We summarize these results as follows: With observations $y_1, y_2, \ldots, y_T$ we compute

$$q_{00} = \sum_{s=1}^{T}\frac{\Delta_{s-1}}{\Delta_s}w_s^2, \quad w_1 = y_1, \quad w_s = y_s - \rho\frac{\Delta_{s-2}}{\Delta_{s-1}}w_{s-1}, \quad s = 2, \ldots, T; \tag{5.39}$$

(5.40)

$$q_{10} = \sum_{s=1}^{T}x_s^2, \quad x_T = \frac{\Delta_{T-1}}{\Delta_T}w_T, \quad x_{T-s} = \frac{\Delta_{T-s-1}}{\Delta_{T-s}}(w_{T-s} - \rho x_{T-s+1}), \quad s = 1, \ldots, T-1;$$

$$q_{20} = \sum_{s=1}^{T}\frac{\Delta_{s-1}}{\Delta_s}h_s^2, \quad h_1 = x_1, \quad h_s = x_s - \rho\frac{\Delta_{s-2}}{\Delta_{s-1}}h_{s-1}, \quad s = 2, \ldots, T. \tag{5.41}$$

5.3. Successive elimination

From the preceding discussion it follows that we have to solve certain linear systems. Let us consider (5.7) in detail, namely, $\boldsymbol{R}\boldsymbol{x} = \boldsymbol{y}$: we have to solve it for $\boldsymbol{x}$ for a given vector of observations $\boldsymbol{y}$, and a given matrix $\boldsymbol{R}$ evaluated during an iterative procedure. The method of successive elimination corresponds to multiplying the system on the left by the matrix $\boldsymbol{F}$ that is lower triangular with diagonal elements 1 so that in the resulting system

$$\boldsymbol{F}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{F}\boldsymbol{y} \tag{5.42}$$

$\boldsymbol{F}\boldsymbol{R}$ is upper triangular. This upper triangular linear systsem is called the "forward solution" of the method of successive (or Gaussian) elimination, or pivotal condensation.

Anderson (1971) gave this procedure in detail for the case of (5.42); see also Anderson (1984), Appendix A, Theorem A.1.2. Using this approach, $\boldsymbol{y}'\boldsymbol{R}^{-2}\boldsymbol{y}$ and $\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{y}$ were evaluated and, for example, the final expression for the former coincides with (5.40). This is so because the Cholesky decomposition is equivalent to the forward part of the method of successive elimination. We summarize these details here for the sake of completeness, and because they provide a practical way to calculate the elements introduced in Section 5.2.

The product $\boldsymbol{FR}$ corresponds to successive left products by elementary matrices $\boldsymbol{F}_j$ so that

$$\boldsymbol{FR} = \boldsymbol{F}_{T-1} \cdots \boldsymbol{F}_2 \boldsymbol{F}_1 \boldsymbol{R}. \tag{5.43}$$

$\boldsymbol{F}_j$ adds to row $j+1$ of the product $\boldsymbol{F}_{j-1} \cdots \boldsymbol{F}_2 \boldsymbol{F}_1 \boldsymbol{R}$ a multiple of its row j so that the product $\boldsymbol{F}_j \boldsymbol{F}_{j-1} \cdots \boldsymbol{F}_2 \boldsymbol{F}_1 \boldsymbol{R}$ has elements equal to 0 in column j below the j,j-th element.

Let $\boldsymbol{F}_s = (f_{ij}^{(s)})$, $s = 1, \ldots, T-1$, $\boldsymbol{F} = (f_{ij})$, $\boldsymbol{Fy} = \boldsymbol{w} = (w_1, \ldots, w_T)'$. Let $r_{jj}^{(j)}$ be the j,j-th element of the product $\boldsymbol{F}_j \boldsymbol{F}_{j-1} \cdots \boldsymbol{F}_2 \boldsymbol{F}_1 \boldsymbol{R}$. Then,

$$r_{11}^{(1)} = 1, \quad r_{jj}^{(j)} = \frac{r_{j-1,j-1}^{(j-1)} - \rho^2}{r_{j-1,j-1}^{(j-1)}}, \qquad j = 2, \ldots, T, \tag{5.44}$$

$$f_{j+1,j}^{(j)} = -\frac{\rho}{r_{jj}^{(j)}}, \qquad j = 1, \ldots, T-1. \tag{5.45}$$

The elements $r_{j+1,j+1}^{(j+1)}$ and $f_{j+1,j}^{(j)}$, $j = 1, \ldots, T-1$, can be computed in sequence. Then compute $\boldsymbol{w}$ as follows:

$$w_1 = y_1, \quad w_j = y_j + f_{j,j-1}^{(j-1)} w_{j-1}, \qquad j = 2, \ldots, T. \tag{5.46}$$

Thus, the elements of $\boldsymbol{w}$ can also be calculated in sequence. Finally calculate

$$x_T = \frac{w_T}{r_{TT}^{(T)}}, \quad x_j = \frac{w_j}{r_{jj}^{(j)}} + f_{j+1,j}^{(j)} x_{j+1}, \qquad j = T-1, \ldots, 1. \tag{5.47}$$

Having calculated $\boldsymbol{x}$ we compute q_{00} and q_{10} using (5.9).

Comparing (5.21) with (5.44) we deduce that

$$r_{jj}^{(j)} = v_{jj} = \frac{\Delta_j}{\Delta_{j-1}}, \tag{5.48}$$

and in fact we are calculating the diagonal elements of $\boldsymbol{V}$ in (5.14) in an (ascending) sequence. Comparing (5.16) with (5.45) we deduce that

$$f_{j+1,j}^{(j)} = -u_{j+1,j}. \tag{5.49}$$

The vector $\boldsymbol{x}$ as given in (5.47) is the "backward solution" of the method of successive elimination or pivotal condensation. In effect, (5.47) can be written as

$$\boldsymbol{x} = \boldsymbol{V}^{-1}\boldsymbol{w} + \begin{pmatrix} 0 & f_{21}^{(1)} & 0 & \ldots & 0 & 0 \\ 0 & 0 & f_{32}^{(2)} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & f_{T,T-1}^{(T-1)} \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix} \boldsymbol{x}, \tag{5.50}$$

where $\boldsymbol{w}$ depends on $y_1, \ldots, y_T$, and this is the form of the backward solution.

5.4. Another recursive procedure

The linear system $\boldsymbol{y} = \boldsymbol{R}\boldsymbol{x}$ can also be solved by repeated substitution, as follows. Using $\boldsymbol{R} = \boldsymbol{I} + \rho\boldsymbol{G}$ we have

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{T-1} \\ x_T \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{T-1} \\ y_T \end{pmatrix} - \begin{pmatrix} \rho x_2 \\ \rho x_1 + \rho x_3 \\ \vdots \\ \rho x_{T-2} + \rho x_T \\ \rho x_{T-1} \end{pmatrix}. \tag{5.51}$$

By repeated substitution we have

$$\begin{aligned} x_1 &= y_1 - \rho x_2 = c_{11} y_1 + b_1 x_2, \\ x_2 &= y_2 - \rho x_1 - \rho x_3 = y_2 + (-\rho) y_1 + (-\rho) x_3 + (-\rho)^2 x_2 \\ &= (1-\rho^2)^{-1}\{(-\rho)y_1 + y_2 + (-\rho)x_3\} = c_{21}y_1 + c_{22}y_2 + b_2 x_3. \end{aligned}$$

In general we have

$$x_t = \sum_{j=1}^{T} c_{tj} y_j + b_t x_{t+1}, \qquad t = 1, \ldots, T-1. \tag{5.52}$$

Then

$$\begin{aligned} x_{t+1} &= y_{t+1} - \rho x_t - \rho x_{t+2} = y_{t+1} - \rho x_{t+2} + (-\rho)\left\{\sum_{j=1}^{t} c_{tj} y_j + b_t x_{t+1}\right\} \\ &= (1+\rho b_t)^{-1}\left\{\sum_{j=1}^{t}(-\rho)c_{tj}y_j + y_{t+1} + (-\rho)x_{t+2}\right\}, \end{aligned}$$

and the recursion for the coefficients is

$$\begin{aligned} c_{t+1,j} &= -\tfrac{\rho}{1+\rho b_t} c_{tj}, & j &= 1, \ldots, t, \\ &= \tfrac{1}{1+\rho b_t}, & j &= t+1, \\ b_t &= -\tfrac{\rho}{1+\rho b_t}, \end{aligned} \tag{5.53}$$

where these expressions hold for $t = 1, \ldots, T-1$, and we either define $b_T = 0$ or take $x_{T+1} = 0$. The resulting system is

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{T-1} \\ x_T \end{pmatrix} = \begin{pmatrix} c_{11} & 0 & \cdots & 0 & 0 \\ c_{21} & c_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ c_{T-1,1} & c_{T-1,2} & \cdots & c_{T-1,T-1} & 0 \\ c_{T1} & c_{T2} & \cdots & c_{T,T-1} & c_{TT} \end{pmatrix}, \tag{5.54}$$

where $c_{11} = 1$, $b_1 = -\rho$. Denoting $\boldsymbol{C} = (c_{ij})$, (5.54) can be written as

$$\boldsymbol{x} = \boldsymbol{C}\boldsymbol{y} + \begin{pmatrix} 0 & b_1 & 0 & \dots & 0 \\ 0 & 0 & b_2 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & b_{T-1} \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix} \boldsymbol{x}, \tag{5.55}$$

which compares with (5.50). In fact,

$$b_t = -\frac{\rho}{v_{tt}} = -\rho\frac{\Delta_{t-1}}{\Delta_t} = f^{(t)}_{t+1,t} = -u_{t+1,t}, \qquad t = 1, \dots, T-1, \tag{5.56}$$

while

$$c_{tj} = \frac{(-\rho)^{t-j}}{v_{jj}} = (-\rho)^{t-j}\frac{\Delta_{j-1}}{\Delta_j}, \qquad j = 1, \dots, t. \tag{5.57}$$

The c_{tj} can be obtained from $\boldsymbol{V}^{-1}\boldsymbol{w}$ in (5.50) by repeated substitutions.

6. Evaluation of Traces in the Time Domain

6.1. Introduction

In Section 6.2 we will evaluate t_{10} and t_{20} by means of series expansions; we now show how all traces t_{jk} used in Section 4, where $j \geq k$, can be expressed in terms of the t_{j0}. In effect, using again that $\boldsymbol{G} = \rho^{-1}(\boldsymbol{R} - \boldsymbol{I})$ we have

$$\begin{aligned} t_{jk} &= \operatorname{tr} \boldsymbol{R}^{-j}\boldsymbol{G}^k = \rho^{-k} \operatorname{tr} \boldsymbol{R}^{-j}(\boldsymbol{R} - \boldsymbol{I})^k \\ &= \rho^{-k} \operatorname{tr} \boldsymbol{R}^{-j} \sum_{s=0}^{k} (-1)^{k-s} \binom{k}{s} \boldsymbol{R}^s = \rho^{-k} \sum_{s=0}^{k} (-1)^{k-s} \binom{k}{s} \operatorname{tr} \boldsymbol{R}^{-(j-s)} \\ &= \rho^{-k} \sum_{s=0}^{k} (-1)^s \binom{k}{s} t_{j-s,0}, \qquad j \geq k. \end{aligned} \tag{6.1}$$

For example

$$t_{11} = \frac{1}{\rho}(T - t_{10}), \quad t_{21} = \frac{1}{\rho}(t_{10} - t_{20}), \quad t_{22} = \frac{1}{\rho^2}(T - 2t_{10} + t_{20}), \tag{6.2}$$

These relations can be used to express the iterative procedures of Section 4 as functions of the various quadratic forms and of the t_{j0}. For example, in Iterative Procedure 1, (4.11) becomes in terms of q_{j0} and t_{j0} only,

$$\left\{\left[\widehat{t}_{20}^{(i-1)}\widehat{q}_{00}^{(i-1)} - \widehat{t}_{10}^{(i-1)}\widehat{q}_{10}^{(i-1)}\right] + T\widehat{q}_{10}^{(i-1)} - \widehat{t}_{10}^{(i-1)}\widehat{q}_{00}^{(i-1)}\right\}\widehat{\rho}^{(i)} = \left[\widehat{t}_{20}^{(i-1)}\widehat{q}_{00}^{(i-1)} - \widehat{t}_{10}^{(i-1)}\widehat{q}_{10}^{(i-1)}\right]\widehat{\rho}^{(i-1)}, \tag{6.3}$$

and in Iterative Procedure 3, equation (4.20) becomes

$$\left\{\left[T - \widehat{t}_{10}^{(i-1)} - \widehat{t}_{20}^{(i-1)}\right]\widehat{q}_{00}^{(i-1)} - \widehat{t}_{10}^{(i-1)}\widehat{q}_{00}^{(i-1)}\right\}\widehat{\rho}^{(i)} = \left\{\left[T - \widehat{t}_{10}^{(i-1)} - \widehat{t}_{20}^{(i-1)}\right]\widehat{q}_{00}^{(i-1)} - T\widehat{q}_{10}^{(i-1)}\right\}\widehat{\rho}^{(i-1)}, \tag{6.4}$$

In Section 6.3 we shall develop a procedure to evaluate t_{11} and t_{22} in the form of some rational expressions; we now show how to express all other needed traces as functions of these two.

From (6.2) we deduce that

$$t_{10} = T - \rho t_{11}, \quad t_{20} = T - 2\rho t_{11} + \rho^2 t_{22}, \tag{6.5}$$

and hence that

$$t_{21} = t_{11} - \rho t_{22}. \tag{6.6}$$

These relations can be used to express the iterative procedures of Section 4. For example, expression (5.3) for Iterative Procedure 1 becomes

$$\left\{\left[\widehat{t}_{11}^{(i-1)} - \widehat{\rho}^{(i-1)}\widehat{t}_{22}^{(i-1)}\right]\left[\widehat{q}_{10}^{(i-1)} - \widehat{q}_{00}^{(i-1)}\right] + \widehat{\rho}^{(i-1)}\widehat{t}_{22}^{(i-1)}\widehat{q}_{10}^{(i-1)}\right\}\widehat{\rho}^{(i)} \tag{6.7}$$
$$= -\left\{T - 2\widehat{\rho}^{(i-1)}\widehat{t}_{11}^{(i-1)} + \left[\widehat{\rho}^{(i-1)}\right]^2\widehat{t}_{22}^{(i-1)}\right\}\left[\widehat{q}_{10}^{(i-1)} - \widehat{q}_{00}^{(i-1)}\right] - \widehat{\rho}^{(i-1)}\left[\widehat{t}_{11}^{(i-1)} - \widehat{\rho}^{(i-1)}\widehat{t}_{22}^{(i-1)}\right]\widehat{q}_{10}^{(i-1)},$$

while expression (5.4) for Iterative Procedure 3 becomes

$$\widehat{\rho}^{(i-1)}\widehat{t}_{22}^{(i-1)}\widehat{q}_{00}^{(i-1)}\widehat{\rho}^{(i)} = \left\{T - \widehat{\rho}^{(i-1)}\widehat{t}_{11}^{(i-1)} + \left[\widehat{\rho}^{(i-1)}\right]^2\widehat{t}_{22}^{(i-1)}\right\}\widehat{q}_{00}^{(i-1)} - T\widehat{q}_{10}^{(i-1)}. \tag{6.8}$$

6.2. Series expressions

For $|\rho| < 1/2$ we have

$$t_{10} = \operatorname{tr} \boldsymbol{R}^{-1} = \operatorname{tr}(\boldsymbol{I} + \rho \boldsymbol{G})^{-1} = \operatorname{tr}\sum_{j=0}^{\infty}(-\rho)^j \boldsymbol{G}^j = \sum_{k=0}^{\infty}\rho^{2k}\operatorname{tr} \boldsymbol{G}^{2k}, \tag{6.9}$$

since tr $\boldsymbol{G}^j = 0$ for j odd. Note that (6.9) converges because the characteristic roots of $\boldsymbol{G}$ are less than 2 in absolute value and $|\rho| \le \frac{1}{2}$. Similarly,

$$t_{20} = \text{tr } \boldsymbol{R}^{-2} = \text{tr}\,(\boldsymbol{I} + \rho \boldsymbol{G})^{-2} = \text{tr} \sum_{j=1}^{\infty} j(-\rho)^{j-1} \boldsymbol{G}^{j-1} = \sum_{k=0}^{\infty} (2k+1)\rho^{2k} \text{ tr } \boldsymbol{G}^{2k}. \tag{6.10}$$

It is shown in Section 7.3 that

$$t_{10} = \frac{T+1}{\sqrt{1-4\rho^2}} - \frac{1}{1-4\rho^2} + 2(T+1) \sum_{g=1}^{\infty} \sum_{k=g(T+1)}^{\infty} \binom{2k}{k - g(T+1)} \rho^{2k}. \tag{6.11}$$

6.3. Rational expressions

To compute t_{11} and t_{22} we consider an expression for $|\boldsymbol{R}|$ and use that

$$\frac{d}{d\rho} \log |\boldsymbol{R}| = \text{tr } \boldsymbol{R}^{-1}\boldsymbol{G} = t_{11}, \tag{6.12}$$

$$\frac{d^2}{d\rho^2} \log |\boldsymbol{R}| = -\text{ tr } \boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{G} = -t_{22}. \tag{6.13}$$

Anderson (1971), Lemma 6.7.9, shows that

$$|\boldsymbol{I} - \theta \boldsymbol{G}| = (1-4\theta^2)^{-\frac{1}{2}} \left(\frac{1}{2}\right)^{T+1} \left\{ \left(1 + \sqrt{1-4\theta^2}\right)^{T+1} - \left(1 - \sqrt{1-4\theta^2}\right)^{T+1} \right\}. \tag{6.14}$$

Hence, we identify $\theta = -\rho$ and use this result directly. Let us denote $a = (1-4\rho^2)^{\frac{1}{2}}$, so that $da/d\rho = -4\rho/a$. Then

$$|\boldsymbol{R}| = |\boldsymbol{I} + \rho \boldsymbol{G}| = \frac{1}{a} \left(\frac{1}{2}\right)^{T+1} \left\{ (1+a)^{T+1} - (1-a)^{T+1} \right\}, \tag{6.15}$$

$$\log |\boldsymbol{R}| = -\log a - (T+1)\log 2 + \log \left\{ (1+a)^{T+1} - (1-a)^{T+1} \right\}, \tag{6.16}$$

$$\begin{aligned} \frac{d}{d\rho} \log |\boldsymbol{R}| &= \frac{4\rho}{a^2} + \frac{(T+1)\frac{4\rho}{a} \left\{ -(1+a)^T - (1-a)^T \right\}}{(1+a)^{T+1} - (1-a)^{T+1}} \\ &= \frac{4\rho}{a^2} - \frac{4(T+1)\rho}{a} \frac{(1+a)^T + (1-a)^T}{(1+a)^{T+1} - (1-a)^{T+1}}. \end{aligned} \tag{6.17}$$

$$(6.18)\qquad \frac{d^2}{d\rho^2}\log|\boldsymbol{R}| = \frac{4a^2+32\rho^2}{a^4} - \frac{4(T+1)(a^2+4\rho^2)}{a^3}\frac{(1+a)^T+(1-a)^T}{(1+a)^{T+1}-(1-a)^{T+1}}$$
$$+\frac{16(T+1)\rho^2}{a^2}\left[\frac{T\{(1+a)^{T-1}-(1-a)^{T-1}\}\{(1+a)^{T+1}-(1-a)^{T+1}\}}{\{(1+a)^{T+1}-(1-a)^{T+1}\}^2}\right.$$
$$\left.-\frac{(T+1)\{(1+a)^T+(1-a)^T\}^2}{\{(1+a)^{T+1}-(1-a)^{T+1}\}^2}\right].$$

Simplifying slightly this last expression we have

$$(6.19)\qquad \frac{d^2}{d\rho^2}\log|\boldsymbol{R}| = \frac{4+16\rho^2}{a^4} - \frac{4(T+1)\{(1+a)^T+(1-a)^T\}}{a^3\{(1+a)^{T+1}-(1-a)^{T+1}\}}$$
$$-\frac{16(T+1)\rho^2}{a^2}\frac{(1+a)^{2T}+(1-a)^{2T}+(4T+8\rho^2)(4\rho^2)^{T-1}}{\{(1+a)^{T+1}-(1-a)^{T+1}\}^2}.$$

6.4. Using the solutions of linear systems

The calculations in Section 5.3 were presented as part of the computations needed for quadratic forms, but can also be used for traces.

From (5.14) we have that $\boldsymbol{R} = \boldsymbol{UVU}'$, so that the "forward solution" (5.42) $\boldsymbol{FRx} = \boldsymbol{Fy} = \boldsymbol{w}$ is $\boldsymbol{VU}'\boldsymbol{x} = \boldsymbol{U}^{-1}\boldsymbol{y} = \boldsymbol{w}$, and simultaneously $\boldsymbol{U}'\boldsymbol{x} = \boldsymbol{V}^{-1}\boldsymbol{U}^{-1}\boldsymbol{y} = \boldsymbol{V}^{-1}\boldsymbol{w}$. Note that $\boldsymbol{F} = \boldsymbol{U}^{-1}$.

To compute $t_{10} = \operatorname{tr} \boldsymbol{R}^{-1}$ we set $\boldsymbol{RX} = \boldsymbol{I}$, where $\boldsymbol{X}$ and $\boldsymbol{I}$ are of orders $T \times T$. In successive elimination the forward solution is $\boldsymbol{FRX} = \boldsymbol{F}$ or $\boldsymbol{VU}'\boldsymbol{X} = \boldsymbol{U}^{-1}$. We get $\boldsymbol{U}^{-1}$ and (diagonal) $\boldsymbol{V}$ by recording the steps of the forward solution of $\boldsymbol{Rx} = \boldsymbol{y}$. Then

$$(6.20)\qquad \begin{aligned} t_{10} &= \operatorname{tr} \boldsymbol{R}^{-1} = \operatorname{tr} \boldsymbol{X} = \operatorname{tr}(\boldsymbol{U}')^{-1}\boldsymbol{V}^{-1}\boldsymbol{U}^{-1} \\ &= \operatorname{tr} \boldsymbol{F}'\boldsymbol{V}^{-1}\boldsymbol{F} = \operatorname{tr} \boldsymbol{V}^{-1}\boldsymbol{FF}' \\ &= 1 + \sum_{s=2}^{T}\frac{1+f_{s1}^2+f_{s2}^2+\cdots+f_{s,s-1}^2}{v_{ss}}, \end{aligned}$$

where $\boldsymbol{F} = (f_{ij})$. From Section 5.3 we have

$$(6.21)\qquad \boldsymbol{F} = \boldsymbol{F}_{T-1}\boldsymbol{F}_{T-2}\cdots\boldsymbol{F}_2\boldsymbol{F}_1,$$

and that the $\boldsymbol{F}_s$ matrices, for $s = 1, \ldots, T-1$, are lower triangular, bidiagonal, and have elements

$$(6.22)\qquad \begin{aligned} f_{ij}^{(s)} &= 1, && i = j, \\ &= -\rho/r_{ss}^{(s)} = -\rho/v_{ss}, && i = s+1, j = s, \\ &= 0, && \text{otherwise.} \end{aligned}$$

Using these results we obtain

$$f_{ij} = f_{i,i-1}^{(i-1)} f_{i-1,i-2}^{(i-2)} \cdots f_{j+1,j}^{(j)}, \quad j = 1, \ldots, T-1;\ i = j+1, \ldots, T. \tag{6.23}$$

In more detail, the elements of $\boldsymbol{F}$ below its main diagonal are

(6.24)

$$\begin{aligned}
&f_{21} = f_{21}^{(1)}, \\
&f_{31} = f_{32}^{(2)} f_{21}^{(1)}, \ f_{32} = f_{32}^{(2)} \\
&f_{41} = f_{43}^{(3)} f_{32}^{(2)} f_{21}^{(1)}, \ f_{42} = f_{43}^{(3)} f_{32}^{(2)}, \ f_{43} = f_{43}^{(3)}, \\
&\vdots \\
&f_{T1} = f_{T,T-1}^{(T-1)} f_{T-1,T-2}^{(T-2)} \cdots f_{21}^{(1)}, \ f_{T2} = f_{T,T-1}^{(T-1)} f_{T-1,T-2}^{(T-2)} \cdots f_{32}^{(2)}, \ldots, f_{T,T-1} = f_{T,T-1}^{(T-1)}
\end{aligned}$$

We next compute t_{20}.

$$\begin{aligned}
t_{20} &= \operatorname{tr} \boldsymbol{R}^{-2} = \operatorname{tr}(\boldsymbol{F}'\boldsymbol{V}^{-1}\boldsymbol{F})(\boldsymbol{F}'\boldsymbol{V}^{-1}\boldsymbol{F}) = \operatorname{tr} \boldsymbol{F}\boldsymbol{F}'\boldsymbol{V}^{-1}\boldsymbol{F}\boldsymbol{F}'\boldsymbol{V}^{-1} \\
&= \operatorname{tr}(\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{F}\boldsymbol{F}'\boldsymbol{V}^{-\frac{1}{2}})(\boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{F}\boldsymbol{F}'\boldsymbol{V}^{-\frac{1}{2}})' = \operatorname{tr} \boldsymbol{H}\boldsymbol{H}' = \sum_{i=1}^{T}\sum_{j=1}^{T} h_{ij}^2,
\end{aligned} \tag{6.25}$$

where $\boldsymbol{H} = \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{F}\boldsymbol{F}'\boldsymbol{V}^{-\frac{1}{2}}$ is symmetric, and we have used the circular property of the trace. The components of $\boldsymbol{F}\boldsymbol{F}'$ are

$$\sum_{s=1}^{T} f_{is} f_{js} = \sum_{s=1}^{\min(i,j)} f_{is} f_{js}, \quad i, j = 1, \ldots, T, \tag{6.26}$$

so that the components of $\boldsymbol{H}$ are

$$h_{ij} = \frac{\sum_{s=1}^{\min(i,j)} f_{is} f_{js}}{\sqrt{v_{ii}}\sqrt{v_{jj}}} = \sum_{s=1}^{\min(i,j)} \frac{f_{is}}{\sqrt{v_{ii}}} \frac{f_{js}}{\sqrt{v_{jj}}}, \quad i, j = 1, \ldots, T, \tag{6.27}$$

and hence

$$\begin{aligned}
t_{20} &= \sum_{i=1}^{T}\sum_{j=1}^{T} \left[\sum_{s=1}^{\min(i,j)} \frac{f_{is}}{\sqrt{v_{ii}}} \frac{f_{js}}{\sqrt{v_{jj}}} \right]^2 \\
&= \sum_{i=1}^{T} \left[\sum_{s=1}^{i} \frac{f_{is}^2}{v_{ii}} \right]^2 + 2 \sum_{\substack{i=1 \\ i<j}}^{T} \sum_{j=1}^{T} \left[\sum_{s=1}^{i} \frac{f_{is}}{\sqrt{v_{ii}}} \frac{f_{js}}{\sqrt{v_{jj}}} \right]^2 \\
&= \sum_{i=1}^{T} \frac{1}{v_{ii}^2} \left[\sum_{s=1}^{i} f_{is}^2 \right]^2 + 2 \sum_{\substack{i=1 \\ i<j}}^{T} \sum_{j=1}^{T} \frac{1}{v_{ii} v_{jj}} \left[\sum_{s=1}^{i} f_{is} f_{js} \right]^2 .
\end{aligned} \tag{6.28}$$

These computations for a given value of ρ can be added to those presented in Section 5.3 to compute q_{00} and q_{10}. We now present the computations needed for q_{00}, q_{10}, q_{20}, t_{10}, and t_{20}, followed by some comments to facilitate the interpretation. Define $\boldsymbol{\Phi} = \boldsymbol{F}\boldsymbol{F}'$.

Starting values ($s = 1$)

$$r_{11}^{(1)} = 1, \tag{6.29}$$

$$f_{jj} = 1, \qquad j = 1, \ldots, T, \tag{6.30}$$

$$t_{10}^{(1)} = 1, \tag{6.31}$$

$$t_{20}^{(1)} = 1, \tag{6.32}$$

$$w_1 = y_1. \tag{6.33}$$

Step s, $s = 2, \ldots, T$

$$r_{ss}^{(s)} = \frac{r_{s-1,s-1}^{(s-1)} - \rho^2}{r_{s-1,s-1}^{(s-1)}}, \tag{6.34}$$

$$f_{s,s-1}^{(s-1)} = -\frac{\rho}{r_{s-1,s-1}^{(s-1)}}, \tag{6.35}$$

$$f_{sj} = f_{s,s-1}^{(s-1)} f_{s-1,j}, \qquad j = 1, \ldots, s-1, \tag{6.36}$$

$$\phi_{sj} = \sum_{k=1}^{j} f_{sk} f_{jk}, \qquad j = 1, \ldots, s, \tag{6.37}$$

$$t_{10}^{(s)} = t_{10}^{(s-1)} + \frac{\phi_{ss}}{r_{ss}^{(s)}}, \tag{6.38}$$

$$t_{20}^{s} = t_{20}^{(s-1)} + \frac{\phi_{ss}^2}{[r_{ss}^{(s)}]^2} + 2\left[\frac{\phi_{s,s-1}^2}{r_{ss}^{(s)} r_{s-1,s-1}^{(s-1)}} + \cdots + \frac{\phi_{s1}^2}{r_{ss}^{(s)} r_{11}^{(1)}}\right], \tag{6.39}$$

$$w_s = y_s + f_{s,s-1}^{(s-1)} w_{s-1}. \tag{6.40}$$

After completing these computations we have

$$t_{10} = t_{10}^{(T)}, \qquad t_{20} = t_{20}^{(T)}, \tag{6.41}$$

$$x_T = \frac{w_T}{r_{TT}^{(T)}} \qquad x_j = \frac{w_j}{r_{jj}^{(j)}} + f_{j+1,j}^{(j)} x_{j+1}, \qquad j = T-1, \ldots, 1, \tag{6.42}$$

$$q_{00} = \sum_{i=1}^{T} y_i x_i, \qquad q_{10} = \sum_{i=1}^{T} x_i^2. \tag{6.43}$$

Finally, to compute q_{20} we have

$$w_1' = x_1, \qquad w_s' = x_s + f_{s,s-1}^{(s-1)} w_{s-1}', \qquad s = 2, \ldots, T, \tag{6.44}$$

$$v_t = \frac{w_T'}{r_{TT}^{(T)}}, \qquad v_j = \frac{w_j'}{r_{jj}^{(j)}} + f_{j+1,j}^{(j)} v_{j+1}, \qquad j = T-1, \ldots, 1, \tag{6.45}$$

$$q_{20} = \sum_{i=1}^{T} x_i v_i. \tag{6.46}$$

Note the following points.

1. Formulas (6.34), (6.35), (6.40), and (6.42) were already given in Section 5.3 to compute q_{00} and q_{10} in (6.43).
2. Similar computations in (6.44) and (6.45) produce q_{20} in (6.46). This was discussed in (5.5) - (5.9).
3. Hence, the superscripts in $r_{ss}^{(s)}$ and $f_{s,s-1}^{(s-1)}$, and the indices in w_s correspond to the calculations being done in sequence.
4. To calculate t_{10} and t_{20} we need to compute the components of $\boldsymbol{F}$ and $\boldsymbol{\Phi} = \boldsymbol{F}\boldsymbol{F}'$. One row of $\boldsymbol{F}$ is computed at each step, namely, for row s, the elements f_{sj} for $j = 1, \ldots, s-1$. Further $f_{ss} = 1$ and $f_{sj} = 0$ for $s < j$. The calculations in (6.36) correspond to the structure of f_{ij} given in (6.23) or (6.24).
5. The components of $\boldsymbol{\Phi} = \boldsymbol{F}\boldsymbol{F}'$ are scalar products of the rows of $\boldsymbol{F}$, and are calculated in (6.37), where the sums can also reach $k = T$ in each case, because for $k > j$ ($j \le s$) at least one of the factors f_{jk} is 0.
6. In $t_{10}^{(s)}$ and $t_{20}^{(s)}$ the superscripts denote partial sums, so that $t_{10}^{(T)} = t_{10}$ and $t_{20}^{(T)} = t_{20}$, which is (6.41).

7. Evaluation of Quadratic Forms and Traces in the Frequency Domain

7.1. Calculation of Fourier coefficients

The Fourier transformation of the operations (3.8) is different from the usual transformation

$$\sqrt{\frac{2}{T}}\sum_{k=1}^{T} y_k \cos\frac{2\pi jk}{T}, \qquad j=0,1,\ldots,\left[\frac{T}{2}\right], \tag{7.1}$$

$$\sqrt{\frac{2}{T}}\sum_{k=1}^{T} y_k \sin\frac{2\pi jk}{T}, \qquad j=1,\ldots,\left[\frac{T-1}{2}\right]. \tag{7.2}$$

Since the transformation (3.8) diagonalizes $\boldsymbol{G}$ and the transformation (7.1) and (7.2) does not diagonalize $\boldsymbol{G}$, the former yields simpler results, as indicated in Section 4.3.

For large T, the fast Fourier transform can be used for efficient computation of (3.8). We write for $j=1,\ldots,T$

$$\begin{aligned} z_j &= \sqrt{\frac{2}{T+1}}\sum_{k=1}^{T} y_k \sin\frac{\pi jk}{T+1} \\ &= \sqrt{\frac{2}{T+1}}\sum_{k=1}^{T+1} y_k \sin\frac{\pi jk}{T+1} \end{aligned} \tag{7.3}$$

for arbitrary y_{T+1} since $\sin\left[\pi jk/(T+1)\right]=0$ for $k=T+1$. Further we have

$$z_j = \sqrt{\frac{1}{2(T+1)}}\sum_{k=1}^{2(T+1)} y_k \sin\frac{2\pi jk}{2(T+1)}, \tag{7.4}$$

where $y_k=-y_{2(T+1)-k}$, $k=T+2,\ldots,2T+1$, and y_{2T+2} is arbitrary. Then (7.4) has the usual form of the sine-transform for $2(T+1)$ observations and the usual computations for the fast Fourier transform are available.

7.2. Evaluation of quadratic forms

We want to calculate

$$\boldsymbol{y}'\boldsymbol{R}^{-(j+1)}\boldsymbol{y} = q_{j0} = \sum_{s=1}^{T}\frac{z_s^2}{(1+\rho d_s)^{j+1}}, \qquad j=0,1,2\ , \tag{7.5}$$

which can be done in a straightforward manner. That $d_s = -d_{T+1-s}$ can be exploited. For example, if T is even

$$(7.6) \qquad q_{00} = \sum_{s=1}^{T} \frac{z_s^2}{1+\rho d_s}$$
$$= \sum_{s=1}^{T/2} \left[\frac{z_s^2}{1+\rho d_s} + \frac{z_{T+1-s}^2}{1-\rho d_s} \right]$$
$$= \sum_{s=1}^{T/2} \frac{(1-\rho d_s) z_s^2 + (1+\rho d_s) z_{T+1-s}^2}{1-\rho^2 d_s^2}$$
$$= \sum_{s=1}^{T/2} \frac{z_s^2 + z_{T+1-s}^2 - \rho d_s (z_s^2 - z_{T+1-s}^2)}{1-\rho^2 d_s^2},$$

while if T is odd, since then $d_{(T+1)/2} = 0$,

$$(7.7) \qquad q_{00} = z_{(T+1)/2}^2 + \sum_{s=1}^{(T-1)/2} \frac{z_s^2 + z_{T+1}^2 - \rho d_s (z_s^2 - z_{T+1-s}^2)}{1-\rho^2 d_s^2}.$$

Using this procedure we have for T even

$$(7.8) \qquad q_{10} = \sum_{s=1}^{T} \frac{z_s^2}{(1+\rho d_s)^2}$$
$$= \sum_{s=1}^{T/2} \left[\frac{z_s^2}{(1+\rho d_s)^2} + \frac{z_{T+1-s}^2}{(1-\rho d_s)^2} \right]$$
$$= \sum_{s=1}^{T/2} \frac{(1-\rho d_s)^2 z_s^2 + (1+\rho d_s)^2 z_{T+1-s}^2}{(1-\rho^2 d_s^2)^2}$$
$$= \sum_{s=1}^{T/2} \frac{(1+\rho^2 d_s^2)(z_s^2 + z_{T+1-s}^2) - 2\rho d_s (z_s^2 - z_{T+1-s}^2)}{(1-\rho^2 d_s^2)^2},$$

while if T is odd,

$$(7.9) \qquad q_{10} = z_{(T+1)/2}^2 + \sum_{s=1}^{(T-1)/2} \frac{(1+\rho^2 d_s^2)(z_s^2 + z_{T+1-s}^2) - 2\rho d_s (z_s^2 - z_{T+1-s}^2)}{(1-\rho^2 d_s^2)^2}.$$

Similarly, for T even

$$(7.10) \qquad q_{20} = \sum_{s=1}^{T} \frac{z_s^2}{(1+\rho d_s)^3}$$

$$
\begin{aligned}
&= \sum_{s=1}^{T/2} \left[\frac{z_s^2}{(1+\rho d_s)^3} + \frac{z_{T+1-s}^2}{(1-\rho d_s)^3} \right] \\
&= \sum_{s=1}^{T/2} \frac{(1-\rho d_s)^3 z_s^2 + (1+\rho d_s)^3 z_{T+1-s}^2}{(1-\rho^2 d_s^2)^3} \\
&= \sum_{s=1}^{T/2} \frac{(1+3\rho^2 d_s^2)(z_s^2 + z_{T+1-s}^2) - (3\rho d_s + \rho^3 d_s^3)(z_s^2 - z_{T+1-s}^2)}{(1-\rho^2 d_s^2)^3},
\end{aligned}
$$

while if T is odd,

$$
q_{20} = z_{(T+1)/2}^2 + \sum_{s=1}^{(T-1)/2} \frac{(1+3\rho^2 d_s^2)(z_s^2 + z_{T+1-s}^2) - (3\rho d_s + \rho^3 d_s^3)(z_s^2 - z_{T+1-s}^2)}{(1-\rho^2 d_s^2)^3}. \tag{7.11}
$$

The series form can be used to obtain

$$
q_{j0} = \sum_{k=0}^{\infty} \frac{\Gamma(j+k+1)}{k!\Gamma(j+1)} (-\rho)^k \sum_{s=1}^{T} d_s^k z_s^2. \tag{7.12}
$$

7.3. Evaluation of traces

Since the characteristic roots of $\boldsymbol{G}$ are

$$
d_s = 2\cos\frac{\pi s}{T+1} = e^{i\frac{\pi s}{T+1}} + e^{-i\frac{\pi s}{T+1}}, \tag{7.13}
$$

the characteristic roots of G^{2k} are

$$
\begin{aligned}
d_s^{2k} &= \left(e^{i\frac{\pi s}{T+1}} + e^{-i\frac{\pi s}{T+1}}\right)^{2k} \\
&= \sum_{j=0}^{2k} \binom{2k}{j} e^{i\frac{\pi s}{T+1}j} e^{-i\frac{\pi s}{T+1}(2k-j)} \\
&= \sum_{j=0}^{2k} \binom{2k}{j} e^{i\frac{2\pi s}{T+1}(k-j)}
\end{aligned} \tag{7.14}
$$

and

$$
\operatorname{tr} \boldsymbol{G}^{2k} = \sum_{j=0}^{2k} \binom{2k}{j} \sum_{s=1}^{T} e^{i\frac{2\pi(k-j)}{T+1}s}. \tag{7.15}
$$

Since

$$(7.16)\qquad \sum_{s=0}^{T} e^{i\frac{2\pi(k-j)}{T+1}s} = T+1 \quad \text{if } (k-j) = 0, \pm(T+1), \pm 2(T+1), \ldots ,$$
$$\qquad = 0 \qquad \text{otherwise,}$$

we have

$$(7.17)\qquad \sum_{s=1}^{T} e^{i\frac{2\pi(k-j)}{T+1}s} = T \qquad \text{if } j = k, k \pm (T+1), k \pm 2(T+1), \ldots ,$$
$$\qquad = -1 \qquad \text{otherwise.}$$

Then

$$(7.18)\qquad \operatorname{tr} \boldsymbol{G}^{2k} = \sum_{j=0}^{2k} \binom{2k}{j}(-1) + (T+1) \sum_{\substack{j=0 \\ j=k,k\pm(T+1),\ldots}}^{2k} \binom{2k}{j}.$$

The first term in tr $\boldsymbol{G}^{2k}$ is -1 times

$$(7.19)\qquad \sum_{j=0}^{2k} \binom{2k}{j} = 2^{2k}.$$

The second term is $T+1$ times

$$(7.20)\qquad \begin{array}{ll} \binom{2k}{k}, & k = 0, 1, \ldots, T, \\ \binom{2k}{k} + 2\binom{2k}{k+(T+1)}, & k = T+1, \ldots, 2T+1, \\ \vdots & \end{array}$$

Then

$$(7.21)\qquad t_{10} = \sum_{k=0}^{\infty} \left\{ \binom{2k}{k}(T+1) - 2^{2k} \right\} \rho^{2k}$$
$$+2(T+1) \sum_{k=T+1}^{\infty} \binom{2k}{k-(T+1)} \rho^{2k} + 2(T+1) \sum_{k=2(T+1)}^{\infty} \binom{2k}{k-2(T+1)} \rho^{2k} + \ldots .$$

The first term in the first sum in (7.21) is $T+1$ times

$$(7.22)\qquad \sum_{k=0}^{\infty} \frac{(2k)!}{(k!)^2} \rho^{2k} = \sum_{k=0}^{\infty} \frac{\Gamma(2k+1)}{k!\Gamma(k+1)} \rho^{2k}$$
$$= \sum_{k=0}^{\infty} \frac{\Gamma(k+\frac{1}{2})\Gamma(k+1)2^{2k}}{k!\Gamma(k+1)\sqrt{\pi}} \rho^{2k}$$
$$= \sum_{k=0}^{\infty} \frac{\Gamma(k+\frac{1}{2})}{k!\Gamma(\frac{1}{2})} (2\rho)^{2k} = \left[1 - (2\rho)^2\right]^{-\frac{1}{2}}.$$

We have used $n! = \Gamma(n+1)$ and the duplication formula for the gamma function

$$(7.23)\qquad \Gamma(2\beta+1) = \frac{\Gamma(\beta+\frac{1}{2})\Gamma(\beta+1)2^{2\beta}}{\sqrt{\pi}}.$$

The first sum in t_{10} is

$$(7.24)\qquad \frac{T+1}{\sqrt{1-4\rho^2}} - \frac{1}{1-4\rho^2},$$

which is a good approximation to t_{10} because the neglected terms are $O(\rho^{2(T+1)})$. Thus we obtain

$$(7.25)\qquad t_{10} = \frac{T+1}{\sqrt{1-4\rho^2}} - \frac{1}{1-4\rho^2} + 2(T+1)\sum_{g=1}^{\infty}\sum_{k=g(T+1)}^{\infty}\binom{2k}{k-g(T+1)}\rho^{2k}.$$

The sum on k in (7.25) can be related to the hypergeometric function as follows: for each fixed $g = 1, 2, \cdots,$

$$\begin{aligned}
(7.26)\ \sum_{k=g(T+1)}^{\infty}\binom{2k}{k-g(T+1)}\rho^{2k}
&= \sum_{k=g(T+1)}^{\infty}\frac{(2k)!}{[k-g(T+1)]![k+g(T+1)]!}\rho^{2k} \\
&= \sum_{h=0}^{\infty}\frac{[2(h+g(T+1))]!}{h![h+2g(T+1)]!}\rho^{2h+2g(T+1)} \\
&= \rho^{2g(T+1)}\sum_{h=0}^{\infty}\frac{\Gamma[2h+2g(T+1)+1]}{h!\Gamma[h+2g(T+1)+1]}\rho^{2h} \\
&= \frac{\rho^{2g(T+1)}2^{2g(T+1)}}{\sqrt{\pi}}\sum_{h=0}^{\infty}\frac{\Gamma[h+g(T+1)+1/2]\Gamma[h+g(T+1)+1]}{h!\Gamma[h+2g(T+1)+1]}(2\rho)^{2h} \\
&= \frac{(2\rho)^{2g(T+1)}}{\sqrt{\pi}}\frac{\Gamma[g(T+1)+1/2]\Gamma[g(T+1)+1]}{\Gamma[2g(T+1)+1]} \\
&\qquad F[g(T+1)+1/2, g(T+1)+1; 2g(T+1)+1; 4\rho^2],
\end{aligned}$$

where we used the definition of the hypergeometric function

$$(7.27)\qquad F(a,b;c;x) = \sum_{j=0}^{\infty}\frac{\Gamma(a+j)}{\Gamma(a)}\frac{\Gamma(b+j)}{\Gamma(b)}\frac{\Gamma(c)}{\Gamma(c+j)}\frac{x^j}{j!}.$$

We conclude that

$$
\begin{aligned}
(7.28) \qquad t_{10} = {} & \frac{T+1}{\sqrt{1-\rho^2}} - \frac{1}{1-4\rho^2} + 2(T+1)\sum_{g=1}^{\infty} \frac{(2\rho)^{2g(T+1)}}{\sqrt{\pi}} \\
& \frac{\Gamma[g(T+1)+1/2]\Gamma[g(T+1)+1]}{\Gamma[2g(T+1)+1]} \\
& F[g(T+1)+1/2, g(T+1)+1; 2g(T+1)+1; 4\rho^2].
\end{aligned}
$$

The argument in Section 7.2 can also be used for traces. In effect,

$$
\begin{aligned}
(7.29) \qquad t_{10} &= \sum_{s=1}^{T} \frac{1}{1+\rho d_s} = 2\sum_{s=1}^{T/2} \frac{1}{1-\rho^2 d_s^2}, \quad \text{for } T \text{ even}, \\
&= 1 + 2\sum_{s=1}^{(T-1)/2} \frac{1}{1-\rho^2 d_s^2}, \quad \text{for } T \text{ odd},
\end{aligned}
$$

$$
\begin{aligned}
(7.30) \qquad t_{20} &= \sum_{s=1}^{T} \frac{1}{(1+\rho d_s)^2} = 2\sum_{s=1}^{T/2} \frac{1+\rho^2 d_s^2}{(1-\rho^2 d_s^2)^2}, \quad \text{for } T \text{ even}, \\
&= 1 + 2\sum_{s=1}^{(T-1)/2} \frac{1+\rho^2 d_s^2}{(1-\rho^2 d_s^2)^2}, \quad \text{for } T \text{ odd}.
\end{aligned}
$$

The expression in (7.24) is an approximation to t_{10}. This value can be obtained by approximating the sum defining t_{10} by a corresponding integral, as was done in Anderson (1971). Besides (7.24) this procedure provides the following approximations:

$$
(7.31) \qquad t_{11} \sim -\frac{1}{\rho}\left\{(T+1)\frac{1-\sqrt{1-1\rho^2}}{\sqrt{1-4\rho^2}} - \frac{4\rho^2}{1-4\rho^2}\right\},
$$

$$
(7.32) \qquad t_{20} \sim \frac{(T+1)\sqrt{1-4\rho^2} - (1+4\rho^2)}{(1-4\rho^2)^2},
$$

$$
(7.33) \qquad t_{21} \sim \frac{-4(T+1)\rho\sqrt{1-4\rho^2} + 8\rho}{(1-4\rho^2)^2},
$$

$$
(7.34) \qquad t_{22} \sim \frac{T+1}{\rho^2}\left\{1 - \frac{1-8\rho^2}{(1-4\rho^2)^{3/2}}\right\} - \frac{4+16\rho^2}{(1-4\rho^2)^2}.
$$

These five approximations also satisfy relations similar to those derived in Section 6.1 among the needed traces, and in particular they satisfy equations similar to (6.2) and (6.5).

8. Other Exact Maximum Likelihood Procedures

In this section we collect some approaches related to the procedures developed in previous sections.

8.1. Calculation of quadratic forms containing the matrix $\boldsymbol{P}$.

We consider an approach that permits the calculation of quadratic forms of the type $\boldsymbol{y}'\boldsymbol{P}^{-j}\boldsymbol{y}$. We illustrate the ideas with the cases $j = 1$ and 2. These quadratic forms can be used to implement iterative procedures in terms of α or to compute quadratic forms in an iterative procedure for ρ by using (2.17).

Let

$$\boldsymbol{L} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad \boldsymbol{a} = \begin{pmatrix} \alpha \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \boldsymbol{B} = \boldsymbol{I} + \alpha\boldsymbol{L}, \ \boldsymbol{Q} = \boldsymbol{B}\boldsymbol{B}'. \tag{8.1}$$

We see that $\boldsymbol{B}$ is nonsingular,

$$\boldsymbol{P} = \boldsymbol{Q} + \boldsymbol{a}\boldsymbol{a}', \tag{8.2}$$

$$\boldsymbol{P}^{-1} = \boldsymbol{Q}^{-1} - \frac{1}{1 + \boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a}}\, \boldsymbol{Q}^{-1}\boldsymbol{a}\boldsymbol{a}'\boldsymbol{Q}^{-1}. \tag{8.3}$$

This is a simple case of a general formula called Woodbury's formula by some authors; see, for example, Phadke and Kedem (1978) and Press (1982).

Calculation of $\boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y}$.

$$\begin{aligned} \boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y} &= \boldsymbol{y}'\boldsymbol{Q}^{-1}\boldsymbol{y} - \frac{1}{1 + \boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a}}\boldsymbol{y}'\boldsymbol{Q}^{-1}\boldsymbol{a}\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{y} \\ &= \boldsymbol{y}'\boldsymbol{B}'^{-1}\boldsymbol{B}^{-1}\boldsymbol{y} - \frac{1}{1 + \boldsymbol{a}'\boldsymbol{B}'^{-1}\boldsymbol{B}^{-1}\boldsymbol{a}}\boldsymbol{y}'\boldsymbol{B}'^{-1}\boldsymbol{B}^{-1}\boldsymbol{a}\boldsymbol{a}'\boldsymbol{B}'^{-1}\boldsymbol{B}^{-1}\boldsymbol{y} \\ &= \boldsymbol{z}'\boldsymbol{z} - \frac{\boldsymbol{z}'\boldsymbol{k}\boldsymbol{k}'\boldsymbol{z}}{1 + \boldsymbol{k}'\boldsymbol{k}} = \boldsymbol{z}'\boldsymbol{z} - \frac{(\boldsymbol{z}'\boldsymbol{k})^2}{1 + \boldsymbol{k}'\boldsymbol{k}}, \end{aligned} \tag{8.4}$$

where

$$B^{-1}y = z, \qquad y = Bz, \tag{8.5}$$

and

$$B^{-1}a = k, \qquad a = Bk. \tag{8.6}$$

We have to find z and k in these two linear systems for given B, y, and a. To solve (8.5) we have

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = (I + \alpha L)z = z + \alpha L z = \begin{pmatrix} z_1 \\ z_2 + \alpha z_1 \\ \vdots \\ z_T + \alpha z_{T-1} \end{pmatrix}, \tag{8.7}$$

and hence,

$$z_1 = y_1, \quad z_j = y_j - \alpha z_{j-1}, \qquad j = 2, \ldots, T. \tag{8.8}$$

These equations can be solved by repeated substitution, giving

$$z_j = \sum_{s=1}^{j} (-\alpha)^{j-s} y_s, \qquad j = 1, \ldots, T. \tag{8.9}$$

To solve (8.6) we have

$$\begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 + \alpha k_1 \\ \vdots \\ k_T + \alpha k_{T-1} \end{pmatrix}, \tag{8.10}$$

and hence,

$$k_1 = \alpha, \quad k_j = -\alpha k_{j-1}, \qquad j = 2, \ldots, T. \tag{8.11}$$

This is solved explicitly as

$$k_j = \alpha(-\alpha)^{j-1} = -(-\alpha)^j, \qquad j = 1, \ldots, T. \tag{8.12}$$

To use in (8.4) we need

$$1 + k'k = 1 + \sum_{j=1}^{T} k_j^2 = 1 + \sum_{j=1}^{T} \alpha^{2j} = 1 + \alpha^2 + \cdots + \alpha^{2T} = \frac{1 - \alpha^{2T+2}}{1 - \alpha^2}. \tag{8.13}$$

We then proceed as follows: Start with

$$z_1 = y_1, \quad S_{11} = z_1^2, \quad S_{21} = \alpha z_1; \tag{8.14}$$

then compute in succession

$$z_j = y_j - \alpha z_{j-1}, \quad S_{1j} = S_{1,j-1} + z_j^2, \quad S_{2j} = S_{2,j-1} - (-\alpha)^j z_j, \qquad j = 2, \ldots, T. \tag{8.15}$$

Then

$$\boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y} = S_{1T} - \frac{1-\alpha^2}{1-\alpha^{2T+2}} S_{2T}^2. \tag{8.16}$$

Calculation of $\boldsymbol{y}'\boldsymbol{P}^{-2}\boldsymbol{y}$.

$$\begin{aligned} \boldsymbol{P}^{-1}\boldsymbol{y} &= \boldsymbol{B}'^{-1}(\boldsymbol{B}^{-1}\boldsymbol{y}) - \frac{1}{1+\boldsymbol{k}'\boldsymbol{k}} \boldsymbol{B}'^{-1}(\boldsymbol{B}^{-1}\boldsymbol{a})(\boldsymbol{B}^{-1}\boldsymbol{a})'(\boldsymbol{B}^{-1}\boldsymbol{y}) \\ &= \boldsymbol{B}'^{-1}\boldsymbol{z} - \frac{1}{1+\boldsymbol{k}'\boldsymbol{k}} \boldsymbol{B}'^{-1}\boldsymbol{k}\boldsymbol{k}'\boldsymbol{z} = \boldsymbol{m} - \frac{\boldsymbol{k}'\boldsymbol{z}}{1+\boldsymbol{k}'\boldsymbol{k}}\boldsymbol{n}, \end{aligned} \tag{8.17}$$

and it suffices to find $\boldsymbol{m}$ and $\boldsymbol{n}$ in the linear systems

$$\boldsymbol{m} = \boldsymbol{B}'^{-1}\boldsymbol{z}, \qquad \boldsymbol{z} = \boldsymbol{B}'\boldsymbol{m}, \tag{8.18}$$

$$\boldsymbol{n} = \boldsymbol{B}'^{-1}\boldsymbol{k}, \qquad \boldsymbol{k} = \boldsymbol{B}'\boldsymbol{n}. \tag{8.19}$$

These systems can be analyzed in the same way as (8.7) and (8.10) to provide the following recursive procedures:

$$m_T = z_T, \qquad m_j = z_j - \alpha m_{j+1}, \qquad j = T-1, \ldots, 1, \tag{8.20}$$

$$n_T = k_T, \qquad n_j = k_j - \alpha n_{j+1}, \qquad j = T-1, \ldots, 1. \tag{8.21}$$

The n_j are given explicitly by

$$n_j = -(-\alpha)^j \frac{1-\alpha^{2(T-j+1)}}{1-\alpha^2}, \qquad j = 1, \ldots, T. \tag{8.22}$$

We proceed as follows: We have $z_1, \ldots, z_T$ available from the calculation of $\boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y}$, and also S_{2T} from (8.15). We then start with

$$m_T = z_T,\ n_T = -(-\alpha)^T,\ S_{3T} = z_T^2,\ S_{4T} = \alpha^{2T},\ S_{5T} = -(-\alpha)^T z_T. \tag{8.23}$$

We then compute in succession

$$m_j = z_j - \alpha m_{j+1},\ n_j = -(-\alpha)^j - \alpha n_{j+1}, \qquad j = T-1, \ldots, 1\ , \tag{8.24}$$

and

$$S_{3j} = S_{3,j-1} + m_j^2,\ S_{4j} = S_{4,j-1} + n_j^2,\ S_{5j} = S_{5,j-1} + m_j n_j, \qquad j = T-1, \ldots, 1\ . \tag{8.25}$$

Then

$$\boldsymbol{y}'\boldsymbol{P}^{-2}\boldsymbol{y} = S_{31} + \left(\frac{1-\alpha^2}{1-\alpha^{2T+2}}\right)^2 S_{2T}^2 S_{41} - 2\frac{1-\alpha^2}{1-\alpha^{2T+2}} S_{2T} S_{51}\ . \tag{8.26}$$

8.2. Estimation using the EM algorithm.

The analysis in the preceding section can be related to the EM algorithm for computing maximum likelihood estimates, as described for example in Dempster, Laird and Rubin (1977).

The generating equations for $y_1, \ldots, y_T$ coming from the MA(1) model (1.1) can be written as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} u_1 + \alpha u_0 \\ u_2 + \alpha u_1 \\ \vdots \\ u_T + \alpha u_{T-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ \alpha & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & \alpha & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix} + u_0 \begin{pmatrix} \alpha \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{8.27}$$

In terms of the notation used in (8.1) we write this as

$$\boldsymbol{y} = (\boldsymbol{I} + \alpha \boldsymbol{L})\boldsymbol{u} + u_0 \boldsymbol{a} = \boldsymbol{B}\boldsymbol{u} + u_0 \boldsymbol{a}, \tag{8.28}$$

which in turn can be written as the transformation

$$\begin{pmatrix} u_0 \\ \boldsymbol{y} \end{pmatrix} = \begin{pmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{a} & \boldsymbol{B} \end{pmatrix} \begin{pmatrix} u_0 \\ \boldsymbol{u} \end{pmatrix}. \tag{8.29}$$

We take $(u_0, \boldsymbol{u}')'$ as $N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{T+1})$. The transformation (8.29) has Jacobian equal to 1, and hence $(u_0, \boldsymbol{y}')'$ is normal with expectation $\boldsymbol{0}$ and covariance matrix

$$\sigma^2 \begin{pmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{a} & \boldsymbol{B} \end{pmatrix} \begin{pmatrix} 1 & \boldsymbol{a}' \\ \boldsymbol{0} & \boldsymbol{B}' \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & \boldsymbol{a}' \\ \boldsymbol{a} & \boldsymbol{a}\boldsymbol{a}' + \boldsymbol{B}\boldsymbol{B}' \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & \boldsymbol{a}' \\ \boldsymbol{a} & \boldsymbol{P} \end{pmatrix}. \tag{8.30}$$

The determinant of this covariance matrix is 1.

To use the EM algorithm we augment the observations $y_1, \ldots, y_T$ by the unobserved u_0 and consider it as a "missing observation." The EM algorithm is an iterative procedure. Given preliminary values of the parameters α and σ^2 we obtain an estimate of u_0 say $u_0^{(1)}$ as the conditional expectation of u_0 given $\boldsymbol{y}$ and preliminary values of α and σ^2. Next we obtain maximum likelihood estimates of α and σ^2 on the basis of $(u_0^{(1)}, \boldsymbol{y}')$. Because α appears only in the exponent of the normal distribution of $(u_0, \boldsymbol{y}')$, this step amounts to minimizing the quadratic form in the exponent of the normal distribution of $(u_0, \boldsymbol{y}')$ and then maximizing the resulting concentrated likelihood with respect to σ^2. However, since the value of σ^2 is irrelevant to maximizing the likelihood with respect to α, one can carry out the iteration with respect to α and after its completion find the estimate of σ^2.

To study the joint density $f(u_0, \boldsymbol{y})$ we use

$$f(u_0, \boldsymbol{y}) = g(u_0|\boldsymbol{y})h(\boldsymbol{y}). \tag{8.31}$$

From the covariance matrix (8.30) we find

$$\mathcal{E}(u_0|\boldsymbol{y}) = \boldsymbol{a}'\boldsymbol{P}^{-1}\boldsymbol{y} = \boldsymbol{a}'\left(\boldsymbol{Q}^{-1} - \frac{1}{1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a}}\boldsymbol{Q}^{-1}\boldsymbol{a}\boldsymbol{a}'\boldsymbol{Q}^{-1}\right)\boldsymbol{y} = \frac{\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{y}}{1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a}}, \tag{8.32}$$

(8.33)

$$\operatorname{Var}(u_0|\boldsymbol{y}) = 1 - \boldsymbol{a}'\boldsymbol{P}^{-1}\boldsymbol{a} = 1 - \boldsymbol{a}'\left(\boldsymbol{Q}^{-1} - \frac{1}{1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a}}\boldsymbol{Q}^{-1}\boldsymbol{a}\boldsymbol{a}'\boldsymbol{Q}^{-1}\right)\boldsymbol{a} = \frac{1}{1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a}},$$

while $\mathcal{E}\boldsymbol{y} = \boldsymbol{0}$, $\operatorname{Var}(\boldsymbol{y}) = \boldsymbol{P}$.

Hence, the exponent in the joint density of $(u_0, \boldsymbol{y}')$ is $-\frac{1}{2}$ times

$$\begin{aligned} &\boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y} + (1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a})\left(u_0 - \frac{1}{1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a}}\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{y}\right)^2 \\ &= (u_0, \boldsymbol{y}')\left\{\begin{pmatrix} 0 & \boldsymbol{0}' \\ \boldsymbol{0} & \boldsymbol{P}^{-1}\end{pmatrix} + \begin{pmatrix} 1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a} & -\boldsymbol{a}'\boldsymbol{Q}^{-1} \\ -\boldsymbol{Q}^{-1}\boldsymbol{a} & (1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a})^{-1}\boldsymbol{Q}^{-1}\boldsymbol{a}\boldsymbol{a}'\boldsymbol{Q}^{-1}\end{pmatrix}\right\}\begin{pmatrix} u_0 \\ \boldsymbol{y}\end{pmatrix}. \end{aligned} \tag{8.34}$$

We now apply the EM algorithm.

E-step. For a given value of α calculate

$$\hat{u}_0 = \mathcal{E}(u_0|\boldsymbol{y}) = \frac{\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{y}}{1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a}}. \tag{8.35}$$

M-step. Minimize with respect to α the quadratic form

$$(8.36)\qquad (\hat{u}_0, \boldsymbol{y}')\left[\begin{pmatrix} 1 & \boldsymbol{0}' \\ \boldsymbol{a} & \boldsymbol{B} \end{pmatrix}\begin{pmatrix} 1 & \boldsymbol{a}' \\ \boldsymbol{0} & \boldsymbol{B}' \end{pmatrix}\right]^{-1}\begin{pmatrix} \hat{u}_0 \\ \boldsymbol{y} \end{pmatrix}$$

$$= (\hat{u}_0, \boldsymbol{y}')\begin{pmatrix} 1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a} & -\boldsymbol{a}'\boldsymbol{Q}^{-1} \\ -\boldsymbol{Q}^{-1}\boldsymbol{a} & \boldsymbol{Q}^{-1} \end{pmatrix}\begin{pmatrix} \hat{u}_0 \\ \boldsymbol{y} \end{pmatrix}$$

$$= \hat{u}_0^2(1+\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{a}) - 2\hat{u}_0\boldsymbol{a}'\boldsymbol{Q}^{-1}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{Q}^{-1}\boldsymbol{y}.$$

With the minimizing value of α repeat the E and M steps.

Since $\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{B}'$, $\boldsymbol{z} = \boldsymbol{B}^{-1}\boldsymbol{y}$, $\boldsymbol{k} = \boldsymbol{B}^{-1}\boldsymbol{a}$ as used in Section 8.1, we have that in this notation (8.35) is

$$(8.37)\qquad \hat{u}_0 = \frac{\boldsymbol{a}'\boldsymbol{B}'^{-1}\boldsymbol{B}^{-1}\boldsymbol{y}}{1+\boldsymbol{a}'\boldsymbol{B}'^{-1}\boldsymbol{B}^{-1}\boldsymbol{a}} = \frac{\boldsymbol{z}'\boldsymbol{k}}{1+\boldsymbol{k}'\boldsymbol{k}},$$

and (8.36) is

$$(8.38)\qquad \hat{u}_0(1+\boldsymbol{k}'\boldsymbol{k}) - 2\hat{u}_0\boldsymbol{z}'\boldsymbol{k} + \boldsymbol{z}'\boldsymbol{z}.$$

If in (8.38) we substitute for $\hat{u}_0$ expression (8.37), we obtain

$$(8.39)\qquad \frac{(\boldsymbol{z}'\boldsymbol{k})^2}{1+\boldsymbol{k}'\boldsymbol{k}} - 2\frac{(\boldsymbol{z}'\boldsymbol{k})^2}{1+\boldsymbol{k}'\boldsymbol{k}} + \boldsymbol{z}'\boldsymbol{z} = \boldsymbol{z}'\boldsymbol{z} - \frac{(\boldsymbol{z}'\boldsymbol{k})^2}{1+\boldsymbol{k}'\boldsymbol{k}} = \boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y}$$

in view of (8.4). Hence, we are minimizing $\boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y}$ with respect to α, but doing the iterations via the EM algorithm.

8.3. Use of the explicit components of the inverse covariance matrix.

As indicated at the beginning of Section 3, the likelihood function can be written as a function of α and σ^2 in terms of the determinant (3.1) and the components (3.4) of $\boldsymbol{P}^{-1}$. In effect,

$$(8.40)\qquad \boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y} = \frac{1}{(1-\alpha^2)(1-\alpha^{2(T+1)})}\Bigg\{\sum_{s=1}^{T} y_s^2(1-\alpha^{2s})(1-\alpha^{2(T-s+1)})$$
$$+2\sum_{s=1}^{T-1}\sum_{t=1}^{T-s} y_s y_{s+t}(-\alpha)^t(1-\alpha^{2s})(1-\alpha^{2(T-s-t+1)})\Bigg\}.$$

Godolphin and de Gooijer (1982) derived from the likelihood function, expressed in these terms, an iterative procedure for α.

8.4. Relation to optimal prediction.

The likelihood function, for example (2.11), can be written in another equivalent form by using (5.25), (5.26), and (5.27). In effect,

$$|\boldsymbol{P}| = |\widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{V}}\widetilde{\boldsymbol{U}}'| = |\widetilde{\boldsymbol{V}}| = \prod_{s=1}^{T} \tilde{v}_{ss}, \tag{8.41}$$

$$\boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y} = (\widetilde{\boldsymbol{U}}^{-1}\boldsymbol{y})'\widetilde{\boldsymbol{V}}^{-1}(\widetilde{\boldsymbol{U}}^{-1}\boldsymbol{y}) = \tilde{\boldsymbol{w}}'\widetilde{\boldsymbol{V}}^{-1}\tilde{\boldsymbol{w}} = \sum_{s=1}^{T} \frac{\tilde{w}_s^2}{\tilde{v}_{ss}}, \tag{8.42}$$

where in analogy to (5.32) we define

$$\tilde{w}_s = y_s - \tilde{u}_{s,s-1}\tilde{w}_{s-1} = y_s - \alpha\frac{1}{\tilde{v}_{s-1,s-1}}\tilde{w}_{s-1}. \tag{8.43}$$

Hence, (2.11) becomes

$$L^*(\alpha, \sigma^2) = (2\pi)^{-T/2}(\sigma^2)^{-T/2}\left(\prod_{s=1}^{T} \tilde{v}_{ss}\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{s=1}^{T}\frac{\tilde{w}_s^2}{\tilde{v}_{ss}}\right\}, \tag{8.44}$$

with $\tilde{v}_{ss} = \tilde{\Delta}_s/\tilde{\Delta}_{s-1}$ defined in (5.26) $\left[\tilde{\Delta}_s = (1-\alpha^{2(s+1)})/(1-\alpha^2) \text{ from } (3.1)\right]$, and $\tilde{w}_s$ defined in (8.43).

This expression can be related to minimum mean square prediction (and hence to Kalman filtering), as several writers have recently emphasized. In effect, we can prove that in our case.

$$\tilde{w}_s = y_s - \mathcal{E}(y_s|y_{s-1}, \ldots, y_1), \qquad s = 2, \ldots, T, \tag{8.45}$$

with $\tilde{w}_1 = 0$, is the error of the optimal prediction of y_s based on $y_{s-1}, \ldots, y_1$, and

$$\sigma^2\tilde{v}_{ss} = \mathrm{Var}(\tilde{w}_s) = \mathcal{E}\left\{\left[y_s - \mathcal{E}(y_s|y_{s-1}, \ldots, y_1)\right]^2 | y_{s-1}, \ldots, y_1\right\}. \tag{8.46}$$

Harvey (1981), while considering the Kalman filtering approach to the problem of estimation in the MA(1) model, gave (8.44), (8.45), and (8.46), and wrote the recursions (in our notation) as

$$\tilde{v}_{ss} = 1 + \frac{\alpha^{2s}}{1 + \alpha^2 + \cdots + \alpha^{2(s-1)}} = \frac{1-\alpha^{2(s+1)}}{1-\alpha^{2s}}, \tag{8.47}$$

(8.48) $$\tilde{w}_s = y_s - \alpha \frac{\tilde{w}_{s-1}}{\tilde{v}_{s-1,s-1}} = y_s + (-\alpha)\frac{1-\alpha^{2(s-1)}}{1-\alpha^{2s}}\tilde{w}_{s-1}.$$

Brockwell and Davies (1987), Section 8.6, gave what corresponds to our analysis in (8.41) - (8.46) of this section, based on a different approach.

We complete the details by using a standard argument in the operation with multivariate normal densities. The likelihood function is obtained by considering the joint normal density of $y_1, \ldots, y_T$ as function of its parameters. A joint density $f(y_1, \ldots, y_T)$ can be written as

(8.49) $$f(y_1, \ldots, y_T) = f(y_T|y_{T-1}, \ldots, y_1)f(y_{T-1}|y_{T-2}, \ldots, y_1)\cdots f(y_2|y_1)f(y_1),$$

that is, a product of conditional (and one marginal) densities. In the multivariate normal case that we consider, all these densities are normal. The expected values can be written as functions of the $(t-1)$-dimensional vectors of covariances

(8.50) $$\left[\mathrm{Cov}(y_t, y_{t-1}), \mathrm{Cov}(y_t, y_{t-2}), \ldots, \mathrm{Cov}(y_t, y_1)\right] = \sigma^2\left[\alpha, 0, \ldots, 0\right], \qquad t = 2, \ldots, T,$$

where we used (2.2), and of the $(t-1)\times(t-1)$ matrix $\boldsymbol{\Sigma}_{t-1}$ that contains the covariances corresponding to the set $y_1, \ldots, y_{t-1}$. Hence,

(8.51) $$\begin{aligned}\mathcal{E}(y_t|y_{t-1}, \ldots, y_1) &= \sigma^2(\alpha, 0, \ldots, 0)'\boldsymbol{\Sigma}_{t-1}^{-1}\begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_1 \end{pmatrix} \\ &= (\alpha, 0, \ldots, 0)'\boldsymbol{P}_{t-1}^{-1}\begin{pmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_1 \end{pmatrix} \\ &= \alpha\sum_{j=1}^{t-1} p_{t-1}^{1j} y_{t-j} = \alpha\sum_{j=1}^{t-1}(-\alpha)^{j-1}\frac{1-\alpha^{2(t-j)}}{1-\alpha^{2t}} y_{t-j} \\ &= -\sum_{j=1}^{t-1}(-\alpha)^j \frac{\tilde{\Delta}_{t-j-1}}{\tilde{\Delta}_{t-1}} y_{t-j},\end{aligned}$$

while for $t = 1$ we take this expected value to be equal to y_1.

Similarly,

(8.52)

$$\operatorname{Var}(y_t|y_{t-1},\ldots,y_1) = \operatorname{Var}(y_t) - \sigma^2(\alpha,0,\ldots,0)'P_{t-1}^{-1}\begin{pmatrix}\alpha\\0\\\vdots\\0\end{pmatrix}$$
$$= \sigma^2(1+\alpha^2) - \sigma^2\alpha^2 p_{t-1}^{11} = \sigma^2(1+\alpha^2) - \sigma^2\alpha^2(1-\alpha^{2(t-1)})/(1-\alpha^{2t})$$
$$= \sigma^2\frac{1-\alpha^{2(t+1)}}{1-\alpha^{2t}} = \sigma^2\frac{\tilde{\Delta}_t}{\tilde{\Delta}_{t-1}} = \sigma^2\tilde{v}_{tt}.$$

Substitution in (8.49) gives

$$L^*(\alpha,\sigma^2) = (2\pi)^{-T/2}\left(\prod_{s=1}^{T}\sigma^2\tilde{v}_{ss}\right)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{s=1}^{T}\frac{1}{\tilde{v}_{ss}}\left[y_s+\sum_{j=1}^{s-1}(-\alpha)^j\frac{\tilde{\Delta}_{s-j-1}}{\tilde{\Delta}_{s-1}}y_{s-j}\right]^2\right\}. \tag{8.53}$$

Comparing with (8.44) we see that it suffices to show that the expression in brackets equals $\tilde{w}_s$. From (8.43), by repeated substitutions, we obtain

$$\begin{aligned}\tilde{w}_s &= y_s - \frac{\alpha}{\tilde{v}_{s-1,s-1}}\tilde{w}_{s-1} = y_s - \frac{1}{\tilde{v}_{s-1,s-1}}y_{s-1} + \alpha^2\frac{1}{\tilde{v}_{s-1,s-1}\tilde{v}_{s-2,s-2}}\tilde{w}_{s-2}\\ &= y_s + (-\alpha)\frac{1}{\tilde{v}_{s-1,s-1}}y_{s-1} + \cdots + (-\alpha)^k\frac{1}{\tilde{v}_{s-1,s-1}\cdots\tilde{v}_{s-k,s-k}}y_{s-k}\\ &\quad + (-\alpha)^{k+1}\frac{1}{\tilde{v}_{s-1,s-1}\cdots\tilde{v}_{s-k-1,s-k-1}}\tilde{w}_{s-k-1},\end{aligned} \tag{8.54}$$

and the result follows because

$$\tilde{v}_{s-1,s-1}\cdots\tilde{v}_{s-k,s-k} = \frac{\tilde{\Delta}_{s-1}}{\tilde{\Delta}_{s-2}}\cdot\frac{\tilde{\Delta}_{s-2}}{\tilde{\Delta}_{s-3}}\cdots\frac{\tilde{\Delta}_{s-k}}{\tilde{\Delta}_{s-k-1}} = \frac{\tilde{\Delta}_{s-1}}{\tilde{\Delta}_{s-k-1}}. \tag{8.55}$$

9. Numbers of Operations Needed to Do the Calculations

Quadratic forms and traces were given (in the frequency domain) in Sections 7.2 and 7.3, respectively, for T even and for T odd. To simplify the analysis in this section we consider the case of T even, since we are interested in orders of magnitude of the numbers of operations.

The traces t_{10} and t_{20} are given in (7.29) and (7.30). Assuming that $d_s = 2\cos\pi s/(T+1)$ is available in the computer, we calculate d_s^2 for $s = 1,\ldots,T/2$ once and for all, and use

that $d_s = -d_{T+1-s}$ for $s = (T/2)+1, \ldots, T$; we also calculate ρ^2 once for each iteration. Then (7.29) requires $T/2$ multiplications $\rho^2 d_s^2$, $T/2$ subtractions $1-\rho^2 d_s^2$, $T/2$ inversions and $T/2$ additions. Formula (7.30) involves additionally $T/2$ additions $1+\rho^2 d_s^2$, $T/2$ multiplications to obtain the square in the denominator, $T/2$ divisions, and $T/2$ additions. For t_{10} and t_{20} we have altogether $2T$ additions or subtractions, T multiplications, and T divisions. See summary table below.

We next consider the calculation of the quadratic forms given in equations (7.6) to (7.11). The calculation of the z_j in (7.4) is about $T \log_2 T$ multiplications and additions, but that is done once and for all. For T even, q_{00} is given in (7.6). The sums $z_s^2 + z_{T+1-s}^2$ and differences $z_s^2 - z_{T+1-s}^2$ are calculated only once. The additional computations for one iteration is T multiplications to obtain ρd_s and then $\rho d_s(z_s^2 - z_{T+1-s}^2)$, $T/2$ additions, $T/2$ divisions, and $T/2$ additions. For T even, q_{10} is given by (7.8). This is additionally T multiplications, $T/2$ subtractions, $T/2$ divisions, and $T/2$ additions. Thus for q_{00} and q_{10} we have $2T$ additions or subtractions, $2T$ multiplications, and T divisions. For T even, q_{20} is given by (7.10). This is additionally $2T$ additions or subtractions, $5T/2$ multiplications, and $T/2$ divisions. Finally, for q_{00}, q_{10}, and q_{20} we have $4T$ additions or subtractions, $9T/2$ multiplications, and $3T/2$ divisions.

The calculations in the time domain were summarized in equations (6.29) to (6.46). To compute q_{00} and q_{10} in (6.43), we use formulas (6.34), (6.35), (6.40), and (6.42), which were also given in Section 5.3. Considering as if we had T steps instead of the $T-1$ actually considered there, they involve $5T$ additions or subtractions, $4T$ multiplications, and $3T$ divisions. To compute q_{20} in (6.46) we use formulas (6.44) and (6.45), that involve additionally $3T$ additions, $3T$ multiplications, and T divisions.

The traces t_{10} and t_{20} are calculated in (6.36), (6.37), (6.38), and (6.39). In (6.36) there are

$$\sum_{s=2}^{T}(s-1) = \sum_{s=1}^{T-1} s = \frac{1}{2}T(T-1) \tag{9.1}$$

multiplications. Then (6.37) involves

$$\begin{aligned} \sum_{s=2}^{T}\sum_{j=1}^{s} j &= \sum_{s=2}^{T}\frac{1}{2}s(s+1) = \frac{1}{2}\sum_{s=1}^{T-1}(s+1)(s+2) = \frac{1}{2}\sum_{s=1}^{T-1}(s^2+3s+2) \\ &= \frac{1}{2}\left\{\frac{(T-1)T(2T-1)}{6} + \frac{3T(T-1)}{2} + 2(T-1)\right\} = \frac{1}{6}(T-1)(T^2+4T+6) \end{aligned} \tag{9.2}$$

multiplications and additions. To obtain $t_{10} = t_{10}^{(T)}$ in (6.38) we need additionally $T-1$ divisions and $T-1$ additions, and to obtain $t_{20} = t_{20}^{(T)}$ in (6.39) we need additionally

$\sum_{s=2}^{T} s = \frac{1}{2}(T+2)(T-1)$ additions, the same number of multiplications, and $\frac{1}{2}T(T-1)$ divisions. Thus t_{10} and t_{20} require altogether $(T-1)(T^2+7T+18)/6$ additions and subtractions, $(T-1)(T^2+10T+12)/6$ multiplications, and $(T-1)(T+2)/2$ divisions.

The number of operations can be summarized as follows:

Quantity Computed	Time Domain +/−	Time Domain ×	Time Domain ÷	Freq. Domain +/−	Freq. Domain ×	Freq. Domain ÷
q_{00}	$4T$	$3T$	$3T$	T	T	$\frac{T}{2}$
q_{10}	T	T	–	T	T	$\frac{T}{2}$
q_{00}, q_{10}	$5T$	$4T$	$3T$	$2T$	$2T$	T
q_{20}	$3T$	$3T$	T	$2T$	$\frac{5T}{2}$	$\frac{T}{2}$
q_{00}, q_{10}, q_{20}	$8T$	$7T$	$4T$	$4T$	$\frac{9T}{2}$	$\frac{3T}{2}$
t_{10}	$\frac{(T-1)(T^2+4T+12)}{6}$	$\frac{(T-1)(T^2+7T+6)}{6}$	$T-1$	T	$\frac{T}{2}$	$\frac{T}{2}$
t_{20}	$\frac{(T-1)(T+2)}{2}$	$\frac{(T-1)(T+2)}{2}$	$\frac{T(T-1)}{2}$	T	$\frac{T}{2}$	$\frac{T}{2}$
t_{10}, t_{20}	$\frac{(T-1)(T^2+7T+18)}{6}$	$\frac{(T-1)(T^2+10T+12)}{6}$	$\frac{(T-1)(T+2)}{2}$	$2T$	T	T

We can compare the different procedures by comparing the number of computations per iteration. The scoring procedures 1 and 3 require the computation of $\hat{q}_{00}^{(i)}$, $\hat{q}_{10}^{(i)}$, $\hat{t}_{10}^{(i)}$, $\hat{t}_{20}^{(i)}$, while the Newton-Raphson procedures require in addition the computation of $\hat{q}_{20}^{(i)}$. It will be seen in the table above that except for the computation of the Fourier coefficients $z_1, \ldots, z_T$ the number of computations carried out in the frequency domain is substantially less than in the time domain. In particular, the number of operations for $\hat{t}_{10}^{(i)}$ and $\hat{t}_{20}^{(i)}$ is of the order $T^3/3$ in the time domain, but of the order $4T$ in the frequency domain. Since the advantage of the frequency domain is in the calculation of the traces, which do not require the Fourier transform of the data, the efficient calculation by any of the procedures is to compute the quadratic forms in the time domain and the traces in the frequency domain.

Of course, counting the number of operations is only one aspect of the evaluation of these methods. Also relevant are the speed of convergence and the behavior in medium-sized samples.

10. Box-Jenkins Procedures

In this section we consider the approach of Box and Jenkins (1976) for computing the quadratic form $\bar{q}_{00}(\alpha) = \boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y}$ and its derivative $d\bar{q}_{00}(\alpha)/d\alpha$ for any given value of α. Box and Jenkins (1976) proposed to estimate α by minimizing $\bar{q}_{00}(\alpha)$; operating with this objective function is different from maximizing the likelihood function or minimizing the concentrated likelihood (2.13) with respect to α because the determinant $|\boldsymbol{P}|$ is ignored. See Box and Jenkins (1976) Chapter 7.

As in Section 8.2, let us consider the transformation from $(u_0, \boldsymbol{u}')'$ which is $N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{T+1})$ to $(u_0, \boldsymbol{y}')'$, defined now by

$$(10.1) \qquad \begin{pmatrix} u_0 \\ \boldsymbol{u} \end{pmatrix} = \boldsymbol{B}^{-1}\begin{pmatrix} u_0 \\ \boldsymbol{y} \end{pmatrix}, \qquad \boldsymbol{B} = \boldsymbol{B}_{T+1} = \boldsymbol{I}_{T+1} + \alpha \boldsymbol{L}_{T+1},$$

where, as in (8.1), $\boldsymbol{L}_{T+1}$ has 1's along the diagonal immediately below its main diagonal and 0's elsewhere. Let

$$(10.2) \qquad \boldsymbol{M} = (\boldsymbol{B}^{-1})'\boldsymbol{B}^{-1} = \begin{pmatrix} m_{00} & \boldsymbol{m}_{01} \\ \boldsymbol{m}_{10} & \boldsymbol{M}_{11} \end{pmatrix},$$

where $\boldsymbol{m}_{10} = \boldsymbol{m}_{01}'$. Then the quadratic form in the exponent of the normal density of $(u_0, \boldsymbol{u}')'$ is $-1/(2\sigma^2)$ times

$$(10.3) \qquad (u_0, \boldsymbol{u}')\begin{pmatrix} u_0 \\ \boldsymbol{u} \end{pmatrix} = (u_0, \boldsymbol{y}')\begin{pmatrix} m_{00} & \boldsymbol{m}_{01} \\ \boldsymbol{m}_{10} & \boldsymbol{M}_{11} \end{pmatrix}\begin{pmatrix} u_0 \\ \boldsymbol{y} \end{pmatrix}$$
$$= m_{00}\left(u_0 + \frac{1}{m_{00}}\boldsymbol{m}_{01}\boldsymbol{y}\right)^2 + \boldsymbol{y}'\left(\boldsymbol{M}_{11} - \frac{1}{m_{00}}\boldsymbol{m}_{10}\boldsymbol{m}_{01}\right)\boldsymbol{y}.$$

Since the Jacobian of the transformation (10.1) is 1, in the normal density of $(u_0, \boldsymbol{u}')'$ we can substitute $\boldsymbol{B}^{-1}(\boldsymbol{u}_0, \boldsymbol{y}')'$ directly to get the normal density of $(u_0, \boldsymbol{y}')'$, and this in turn can be expressed as the product of the marginal density of $\boldsymbol{y}$ times the conditional density of u_0 given $\boldsymbol{y}$. The quadratic form in the exponent of the marginal normal density of $\boldsymbol{y}$ is $-1/(2\sigma^2)$ times

$$(10.4) \qquad \bar{q}_{00}(\alpha) = \boldsymbol{y}'\left(\boldsymbol{M}_{11} - \frac{1}{m_{00}}\boldsymbol{m}_{10}\boldsymbol{m}_{01}\right)\boldsymbol{y},$$

and the quadratic form in the exponent of the conditional normal density of u_0 given $\boldsymbol{y}$ is $-m_{00}\left(u_0 + \frac{1}{m_{00}}\boldsymbol{m}_{01}\boldsymbol{y}\right)^2/(2\sigma^2)$. Thus

$$(10.5) \qquad \mathcal{E}(u_0|\boldsymbol{y}) = -\frac{1}{m_{00}}\boldsymbol{m}_{01}\boldsymbol{y}, \qquad \mathcal{E}\left[\begin{pmatrix} u_0 \\ \boldsymbol{y} \end{pmatrix} \Big| \, \boldsymbol{y}\right] = \begin{pmatrix} -\frac{1}{m_{00}}\boldsymbol{m}_{01} \\ \boldsymbol{I} \end{pmatrix}\boldsymbol{y}$$

because $\boldsymbol{y} = \mathcal{E}(\boldsymbol{y}|\boldsymbol{y})$, and hence

(10.6)

$$\left\{\mathcal{E}\left[\begin{pmatrix} u_0 \\ \boldsymbol{u} \end{pmatrix} \middle| \boldsymbol{y}\right]\right\}' \left\{\mathcal{E}\left[\begin{pmatrix} u_0 \\ \boldsymbol{u} \end{pmatrix} \middle| \boldsymbol{y}\right]\right\} = \boldsymbol{y}' \left(-\frac{1}{m_{00}}\boldsymbol{m}_{10}, \boldsymbol{I}\right) (\boldsymbol{B}^{-1})' \boldsymbol{B}^{-1} \begin{pmatrix} -\frac{1}{m_{00}}\boldsymbol{m}_{01} \\ \boldsymbol{I} \end{pmatrix} \boldsymbol{y}$$
$$= \boldsymbol{y}' \left(\boldsymbol{M}_{11} - \frac{1}{m_{00}}\boldsymbol{m}_{10}\boldsymbol{m}_{01}\right) \boldsymbol{y} = \bar{q}_{00}(\alpha).$$

In conclusion, we have shown that

$$\bar{q}_{00}(\alpha) = \boldsymbol{y}'\boldsymbol{P}^{-1}\boldsymbol{y} = \boldsymbol{y}' \left(\boldsymbol{M}_{11} - \frac{1}{m_{00}}\boldsymbol{m}_{10}\boldsymbol{m}_{01}\right) \boldsymbol{y} = \sum_{t=0}^{T} \left[\mathcal{E}(u_t|\boldsymbol{y})\right]^2. \tag{10.7}$$

To compute $\mathcal{E}(u_t|\boldsymbol{y})$ for $t = 0, 1, \ldots, T$ we use the process u_t introduced in (2.1) and a process of independent normal $(0, \sigma^2)$ random variables for which

$$y_t = v_t + \alpha v_{t+1}, \tag{10.8}$$

that is, has the "time" reversed. Box and Jenkins (1976), Chapter 6, call this the "backward" form of the process.

From model (2.1) we have for a given value of α the recursive relations

$$\mathcal{E}(u_t|\boldsymbol{y}) = y_t - \alpha\mathcal{E}(u_{t-1}|\boldsymbol{y}), \qquad t = 1, \ldots, T, \tag{10.9}$$

which would provide all needed conditional expectations if $\mathcal{E}(u_0|\boldsymbol{y})$ were known. It turns out that we can obtain $\mathcal{E}(u_0|\boldsymbol{y})$ from a recursive relation derived from (10.8), namely

$$\mathcal{E}(v_t|\boldsymbol{y}) = y_t - \alpha\mathcal{E}(v_{t+1}|\boldsymbol{y}), \qquad t = T, \ldots, 1, \tag{10.10}$$

if we make the additional assumption that for some sufficiently large T^*, $1 \le T^* \le T$, $\mathcal{E}(v_{T^*}|\boldsymbol{y}) = 0$. We note that $\mathcal{E}(u_t|\boldsymbol{y}) = 0$ for $t = 0, -1, -2, \ldots$ and for $t = T+1, T+2, \ldots$; similarly, $\mathcal{E}(v_t|\boldsymbol{y}) = 0$ for $t = 0, -1, -2, \ldots$ and for $t = T+2, T+3, \ldots$.

Starting with $\mathcal{E}(v_{T^*}|\boldsymbol{y}) = 0$ and using (10.10), we obtain $\mathcal{E}(v_{T^*-1}|\boldsymbol{y}), \ldots, \mathcal{E}(v_1|\boldsymbol{y})$. For $t = 0$ (10.10) yields $0 = \mathcal{E}(y_0|\boldsymbol{y}) = \alpha\mathcal{E}(v_1|\boldsymbol{y})$. Then use of (10.9) for $t = 0$ yields $\mathcal{E}(u_0|\boldsymbol{y}) = \mathcal{E}(y_0|\boldsymbol{y}) = \alpha\mathcal{E}(v_1|\boldsymbol{y})$, which is the desired starting value.

If more accuracy is needed, one can obtain $\mathcal{E}(y_{T+1}|\boldsymbol{y}) = \alpha\mathcal{E}(u_T|\boldsymbol{y})$ and $\mathcal{E}(v_{T+1}|\boldsymbol{y}) = \mathcal{E}(y_{T+1}|\boldsymbol{y})$ to begin another round of recursions.

Finally we are in a position to compute $\bar{q}_{00}(\alpha) = \sum_{t=0}^{T} \left[\mathcal{E}(u_t|\boldsymbol{y})\right]^2$ for any given value of α. The analysis of $\bar{q}_{00}(\alpha)$ is illustrated in Box and Jenkins (1976), Section 7.1, where it is denoted by $S(\boldsymbol{\theta})$ in general.

Suppose that α_0 is an initial value of α, and let $\mathcal{E}(u_t|\boldsymbol{y},\alpha_0)$ denote the value of the conditional expectation of u_t given $\boldsymbol{y}$ calculated for this value α_0. For any α we can approximate $\mathcal{E}(u_t|\boldsymbol{y},\alpha)$ as

$$\mathcal{E}(u_t|\boldsymbol{y},\alpha) \sim \mathcal{E}(u_t|\boldsymbol{y},\alpha_0) + \left.\frac{d\mathcal{E}(u_t|\boldsymbol{y},\alpha)}{d\alpha}\right|_{\alpha=\alpha_0}(\alpha-\alpha_0), \tag{10.11}$$

and $\bar{q}_{00}(\alpha)$ as

$$\bar{q}_{00}(\alpha) \sim \sum_{t=0}^{T}\left\{\mathcal{E}(u_t|\boldsymbol{y},\alpha_0) + \left.\frac{d\mathcal{E}(u_t|\boldsymbol{y},\alpha)}{d\alpha}\right|_{\alpha=\alpha_0}(\alpha-\alpha_0)\right\}^2. \tag{10.12}$$

Minimization of (10.12) with respect to α occurs at

$$\alpha-\alpha_0 = \frac{\sum_{t=0}^{T}\mathcal{E}(u_t|\boldsymbol{y},\alpha_0)\left.\frac{d\mathcal{E}(u_t|\boldsymbol{y},\alpha)}{d\alpha}\right|_{\alpha=\alpha_0}}{\sum_{t=0}^{T}\left[\left.\frac{d\mathcal{E}(u_t|\boldsymbol{y},\alpha)}{d\alpha}\right|_{\alpha=\alpha_0}\right]^2}. \tag{10.13}$$

From (10.9) and (10.10) we obtain

$$\frac{d\mathcal{E}(u_t|\boldsymbol{y})}{d\alpha} = -\mathcal{E}(u_{t-1}|\boldsymbol{y}) - \alpha\frac{d\mathcal{E}(u_{t-1}|\boldsymbol{y})}{d\alpha}, \qquad t=1,\ldots,T, \tag{10.14}$$

$$\frac{d\mathcal{E}(v_t|\boldsymbol{y})}{d\alpha} = -\mathcal{E}(v_{t+1}|\boldsymbol{y}) - \alpha\frac{d\mathcal{E}(v_{t+1}|\boldsymbol{y})}{d\alpha}, \qquad t=1,\ldots,T, \tag{10.15}$$

since $y_t,\ t=1,\ldots,T$ does not depend on α. Further, we need

$$\frac{d\mathcal{E}(u_0|\boldsymbol{y})}{d\alpha} = \frac{d\mathcal{E}(y_0|\boldsymbol{y})}{d\alpha} \tag{10.16}$$

since $\mathcal{E}(u_{-1}|\boldsymbol{y}) = \mathcal{E}(u_{-2}|\boldsymbol{y}) = \cdots = 0$. To calculate (10.16) we use an approximation obtained by calculating (10.15) recursively from some T^* replacing $d\mathcal{E}(v_{T^*}|\boldsymbol{y})/d\alpha$ by 0. This leads to

$$0 = \frac{d\mathcal{E}(v_0|\boldsymbol{y})}{d\alpha} = \frac{d\mathcal{E}(y_0|\boldsymbol{y})}{d\alpha} - \alpha\frac{d\mathcal{E}(v_1|\boldsymbol{y})}{d\alpha} - \mathcal{E}(v_1|\boldsymbol{y}), \tag{10.17}$$

which is solved for $d\mathcal{E}(y_0|\boldsymbol{y})/d\alpha$. Thus we obtain the constituents of (10.13).

The Taylor-series expansion is considered by Box and Jenkins (1976) in Section 7.2.

As in Section 9 we can now count the number of operations needed to do the calculations. These can be summarized as follows

Quantity Computed	$+/-$	$\times$
$\mathcal{E}(v_t\|\boldsymbol{y})$	$T+1$	$T+1$
$\mathcal{E}(u_t\|\boldsymbol{y})$	$T+1$	$T+1$
$\bar{q}_{00}(\alpha)$	T	$T+1$
$\dfrac{d\mathcal{E}(v_t\|\boldsymbol{y})}{d\alpha}$	$T+1$	$T+1$
$\dfrac{d\mathcal{E}(u_t\|\boldsymbol{y})}{d\alpha}$	$T+1$	$T+1$
$\dfrac{d\bar{q}_{00}(\alpha)}{d\alpha}$	$T+1$	$T+1$
Total	$6T+5$	$6T+6$

It should be emphasized that the minimization of $\bar{q}_{00}(\alpha)$ is not the same as the maximization of the loglikelihood because of the factor $\log|\boldsymbol{P}| = \log(1-\alpha^{2T+2})/(1-\alpha^2)$. Even for T so large that α^{2T+2} is negligible the term $\log(1-\alpha^2) \sim -\alpha^2$ may not be small enough to ignore. Each procedure studied in detail in this paper is exactly maximum likelihood in the sense that the iteration is meant to converge to a local maximum. Therefore, these iterative maximum likelihood procedures are not directly comparable to the Box-Jenkins procedures.

References

Ali, M. M. (1977). Analysis of autoregressive-moving average models: estimation and prediction, *Biometrika,* **64**, 535–545.

Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, New York: John Wiley and Sons.

Anderson, T. W. (1975). Maximum likelihood estimation of parameters of an autoregressive process with moving average residuals and other covariance matrices with linear structure, *Annals of Statistics,* **3**, 1283–1304.

Anderson, T. W. (1977). Estimation for autoregressive moving average models in the time and frequency domains, *Annals of Statistics,* **5**, 842–865.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, Second Edition, New York: John Wiley and Sons.

Anderson, T. W. (1987). Discussion of T. P. Speed, What is an analysis of variance? *Annals of Statistics,* **15**, 911–913.

Anderson, T. W., and R. P. Ment- (1980). On the structure of the likelihood function of autoregressive and moving average models, *Journal of Time Series Analysis,* **1**, 83–94.

Anderson, T. W., and Akimichi Takemura (1986). Why do noninvertible estimated moving averages occur? *Journal of Time Series Analysis,* **7**, 235–254.

Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive moving average process, *Biometrika,* **66**, 59–65.

BMDP (1985). *BMDP Statistical Software*, Los Angeles, California.

Box, G. W. P., and G. M. Jenkins (1976). *Time Series Analysis, Forecasting and Control*, Revised Edition, San Francisco: Holden-Day.

Brockwell, P. J., and R. A. Davis (1987). *Time Series: Theory and Methods*, New York: Springer-Verlag.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B,* **39**, 1–38.

Dent, W. (1977). Computation of the exact likelihood function of an ARIMA process, *Journal of Statistical Computation and Simulation.* **5**, 193–206.

Godolphin, E. J., and J. G. de Gooijer (1982). On the maximum likelihood estimation of the parameters of Gaussian moving average process, *Biometrika,* **69**, 443–451.

Golub, G. H., and C. F. Van Loan (1983). *Matrix Computations*, Oxford: North Oxford Academic.

Graybill, F. A. (1969). *Introduction to Matrices with Applications in Statistics*, California: Wadsworth.

Harvey, A. C. (1981). *Time Series Models*, Oxford: Philip Allan.

Hasza, D. F. (1980). A note on maximum likelihood estimation for the first-order autoregressive process, *Communications in Statistics,* **A9**, 1411–1415.

Hillmer, S. G., and G. C. Tiao (1979). Likelihood function of stationary multiple autoregressive moving average models, *Journal of the American Statistical Association,* **74**, 652–661.

IMSL (1979). *The IMSL Library*, Houston, Texas.

Newbold, P. (1974). The exact likelihood function for a mixed autoregressive-moving average process, *Biometrika,* **61**, 423–426.

Nicholls, D. F., and A. D. Hall (1979). The exact likelihood function of multivariate autoregressives moving average models, *Biometrika,* **66**, 259–264.

Osborn, D. R. (1977). Exact and approximate maximum likelihood estimators for vector moving average processes, *Journal of the Royal Statistical Society, Series B,* **39**, 114–118.

Phadke, M. S., and G. Kedem (1978). Computation of the exact likelihood function of multivariate moving average models, *Biometrika,* **65**, 511–520.

Press, S. James (1982). *Applied Multivariate Analysis*, Second Edition, Florida: R. E. Krieger.

Shaman, P. (1969). On the inverse of the covariance matrix of a first-order moving average, *Biometrika,* **56**, 595–600.

Speed, T. P. (1987). What is an analysis of variance? *Annals of Statistics,* **15**, 885–910.